Plant Glutamine Phenylpyruvate Transaminase Gene and Transgenic Plants Carrying Same

US 20100263090 A1

Publication number	US20100263090 A1
Publication type	Application
Application number	US 12/551,320
Publication date	Oct 14, 2010
Filing date	Aug 31, 2009
Priority date	Aug 29, 2008
Also published as	WO2011025515A1
Inventors	Pat J. Unkefer, Penelope S. Anderson, Thomas J. Knight
Original Assignee	Los Alamos National Security, Llc
Export Citation	BiBTeX, EndNote, RefMan

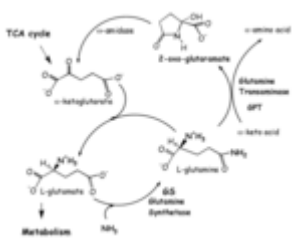
Classifications (9), Legal Events (3)

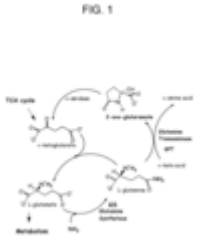
External Links: USPTO, USPTO Assignment, Espacenet

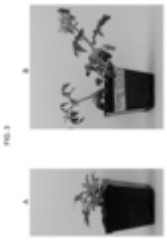
ABSTRACT

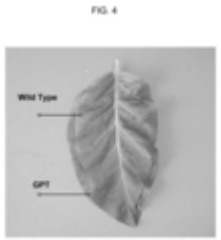
The invention relates to transgenic plants exhibiting enhanced growth rates, seed and fruit yields, and overall biomass yields, as well as methods for generating growth-enhanced transgenic plants. In one embodiment, transgenic plants engineered to over-express glutamine phenylpyruvate transaminase (GPT) are provided.

IMAGES (5)









DESCRIPTION

RELATED APPLICATIONS

This application claims priority to U.S. Provisional Application No. 61/190,581

CLAIMS (21)

1. A transgenic plant comprising a GPT transgene operably linked to a plant promoter.

filed Aug. 29, 2008.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with government support under Contract No. W-7405-ENG-36 awarded by the United States Department of Energy to The Regents of The University of California, and Contract No. DE-AC52-06NA25396, awarded by the United States Department of Energy to Los Alamos National Security, LLC. The government has certain rights in this invention.

BACKGROUND OF THE INVENTION

As the human population increases worldwide, and available farmland continues to be destroyed or otherwise compromised, the need for more effective and sustainable agriculture systems is of paramount interest to the human race. Improving crop yields, protein content, and plant growth rates represent major objectives in the development of agriculture systems that can more effectively respond to the challenges presented.

In recent years, the importance of improved crop production technologies has only increased as yields for many well-developed crops have tended to plateau. Many agricultural activities are time sensitive, with costs and returns being dependent upon rapid turnover of crops or upon time to market. Therefore, rapid plant growth is an economically important goal for many agricultural businesses that involve high-value crops such as grains, vegetables, berries and other fruits.

Genetic engineering has and continues to play an increasingly important yet controversial role in the development of sustainable agriculture technologies. A large number of genetically modified plants and related technologies have been developed in recent years, many of which are in widespread use today (Factsheet: *Genetically Modified Crops in the United States*, Pew Initiative on Food and Biotechnology, August 2004, http://pewagbiotech.org/resources/factsheets). The adoption of transgenic plant varieties is now very substantial and is on the rise, with approximately 250 million acres planted with transgenic plants in 2006.

2. The transgenic plant of claim 1, wherein the GPT transgene encodes a polypeptide having an amino acid sequence selected from the group consisting of (a) SEQ ID NO: 2; SEQ ID NO: 9; SEQ ID NO: 15, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO 24, SEQ ID NO: 30, SEQ ID NO:31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35 and SEQ ID NO: 36, and (b) an amino acid sequence that is at least 75% identical to any one of SEQ ID NO: 2; SEQ ID NO: 9; SEQ ID NO: 15, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO 24, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35 and SEQ ID NO: 36 and has GPT activity.

3. The transgenic plant according to claim 1, wherein the GPT transgene is incorporated into the genome of the plant.

4. The transgenic plant of claim 5, further defined as a monocotyledonous plant.

5. The transgenic plant of claim 5, further defined as a dicotyledonous plant.

6. A progeny of any generation of the transgenic plant of claim 3, wherein said progeny comprises said GPT transgene.

7. A seed of any generation of the transgenic plant of claim 3, wherein said seed comprises said GPT transgene.

8. The transgenic plant of claim 3 which displays an enhanced growth rate when compared to an analogous wild-type or untransformed plant.

9. The transgenic plant of claim 3 which displays increased biomass yield when compared to an analogous wild-type or untransformed plant.

10. The transgenic plant of claim 3 which displays increased seed yield when compared to an analogous wild-type or untransformed plant.

11. The transgenic plant of claim 3 which displays increased flower or flower bud yield when compared to an analogous wild-type or untransformed plant.

While acceptance of transgenic plant technologies may be gradually increasing, particularly in the United States, Canada and Australia, many regions of the World remain slow to adopt genetically modified plants in agriculture, notably Europe. Therefore, consonant with pursuing the objectives of responsible and sustainable agriculture, there is a strong interest in the development of genetically engineered plants that do not introduce toxins or other potentially problematic substances into plants and/or the environment. There is also a strong interest in minimizing the cost of achieving objectives such as improving herbicide tolerance, pest and disease resistance, and overall crop yields. Accordingly, there remains a need for transgenic plants that can meet these objectives.

The goal of rapid plant growth has been pursued through numerous studies of various plant regulatory systems, many of which remain incompletely understood. In particular, the plant regulatory mechanisms that coordinate carbon and nitrogen metabolism are not fully elucidated. These regulatory mechanisms are presumed to have a fundamental impact on plant growth and development.

The metabolism of carbon and nitrogen in photosynthetic organisms must be regulated in a coordinated manner to assure efficient use of plant resources and energy. Current understanding of carbon and nitrogen metabolism includes details of certain steps and metabolic pathways which are subsystems of larger systems. In photosynthetic organisms, carbon metabolism begins with CO_2 fixation, which proceeds via two major processes, termed C-3 and C-4 metabolism. In plants with C-3 metabolism, the enzyme ribulose bisphosphate carboxylase (RuBisCo) catalyzes the combination of CO_2 with ribulose bisphosphate to produce 3-phosphoglycerate, a three carbon compound (C-3) that the plant uses to synthesize carbon-containing compounds. In plants with C-4 metabolism, CO_2 is combined with phosphoenol pyruvate to form acids containing four carbons (C-4), in a reaction catalyzed by the enzyme phosphoenol pyruvate carboxylase. The acids are transferred to bundle sheath cells, where they are decarboxylated to release CO_2, which is then combined with ribulose bisphosphate in the same reaction employed by C-3 plants.

Numerous studies have found that various metabolites are important in plant

12. The transgenic plant of claim 3 which displays increased fruit or pod yield when compared to an analogous wild-type or untransformed plant.

13. The transgenic plant of claim 3 which displays larger leaves when compared to an analogous wild-type or untransformed plant.

14. The transgenic plant of claim 3 which displays increased GPT activity when compared to an analogous wild-type or untransformed plant.

15. The transgenic plant of claim 3 which displays increased GS activity when compared to an analogous wild-type or untransformed plant.

16. The transgenic plant of claim 3 which displays increased 2-oxoglutaramate levels when compared to an analogous wild-type or untransformed plant.

17. The transgenic plant of claim 3 which displays increased nitrogen use efficiency when compared to an analogous wild-type or untransformed plant.

18. A method for producing a plant having enhanced growth properties relative to an analogous wild type or untransformed plant, comprising:

(a) introducing into and expressing in a plant a GPT transgene to produce a biologically active GPT protein; and,

(b) selecting a plant having an increased growth characteristic relative to a plant of the same species that does not comprise a GPT transgene.

19. The method according to claim 17, wherein the increased growth characteristic is selected from the group consisting of increased biomass, earlier flowering, earlier budding, increased plant height, increased flowering, increased budding, larger leaves, increased fruit or pod yield and increased seed yield.

20. A method of producing a plant having increased nitrogen use efficiency

regulation of nitrogen metabolism. These compounds include the organic acid malate and the amino acids glutamate and glutamine. Nitrogen is assimilated by photosynthetic organisms via the action of the enzyme glutamine synthetase (GS) which catalyzes the combination of ammonia with glutamate to form glutamine. GS plays a key role in the assimilation of nitrogen in plants by catalyzing the addition of ammonium to glutamate to form glutamine in an ATP-dependent reaction (Miflin and Habash, 2002, Journal of Experimental Botany, Vol. 53, No. 370, pp. 979-987). GS also reassimilates ammonia released as a result of photorespiration and the breakdown of proteins and nitrogen transport compounds. GS enzymes may be divided into two general classes, one representing the cytoplasmic form (GS1) and the other representing the plastidic (i.e., chloroplastic) form (GS2).

relative to an analogous wild type or untransformed plant, comprising:

(a) introducing and expressing a GPT transgene into the plant;

(b) selecting a plant having an increased nitrogen use efficiency relative to a plant of the same species that does not comprise a GPT transgene.

21. The method according to claim 18, 19 or 20, further comprising propagating a plant from the seed so selected and harvesting a seed therefrom.

Previous work has demonstrated that increased expression levels of GS1 result in increased levels of GS activity and plant growth, although reports are inconsistent. For example, Fuentes et al. reported that CaMV S35 promoter—driven overexpression of Alfalfa GS1 (cytoplasmic form) in tobacco resulted in increased levels of GS expression and GS activity in leaf tissue, increased growth under nitrogen starvation, but no effect on growth under optimal nitrogen fertilization conditions (Fuentes et al., 2001, J. Exp. Botany 52: 1071-81). Temple et al. reported that transgenic tobacco plants overexpressing the full length Alfalfa GS1 coding sequence contained greatly elevated levels of GS transcript, and GS polypeptide which assembled into active enzyme, but did not report phenotypic effects on growth (Temple et al., 1993, Molecular and General Genetics 236: 315-325). Corruzi et al. have reported that transgenic tobacco overexpressing a pea cytosolic GS1 transgene under the control of the CaMV S35 promoter show increased GS activity, increased cytosolic GS protein, and improved growth characteristics (U.S. Pat. No. 6,107,547). Unkefer et al. have more recently reported that transgenic tobacco plants overexpressing the Alfalfa GS1 in foliar tissues, which had been screened for increased leaf-to-root GS activity following genetic segregation by selfing to achieve increased GS1 transgene copy number, were found to produce increased 2-hydroxy-5-oxoproline levels in their foliar portions, which was found to lead to markedly increased growth rates over wildtype tobacco plants (see, U.S. Pat. Nos. 6,555,500; 6,593,275; and 6,831,040).

Unkefer et al. have further described the use of 2-hydroxy-5-oxoproline (also known as 2-oxoglutaramate) to improve plant growth (U.S. Pat. Nos. 6,555,500; 6,593,275; 6,831,040). In particular, Unkefer et al. disclose that increased concentrations of 2-hydroxy-5-oxoproline in foliar tissues (relative to root tissues) triggers a cascade of events that result in increased plant growth characteristics. Unkefer et al. describe methods by which the foliar concentration of 2-hydroxy-5-oxoproline may be increased in order to trigger increased plant growth characteristics, specifically, by applying a solution of 2-hydroxy-5-oxoproline directly to the foliar portions of the plant and over-expressing glutamine synthetase preferentially in leaf tissues.

A number of transaminase and hydrolyase enzymes known to be involved in the synthesis of 2-hydroxy-5-oxoproline in

animals have been identified in animal liver and kidney tissues (Cooper and Meister, 1977, CRC Critical Reviews in Biochemistry, pages 281-303; Meister, 1952, J. Biochem. 197: 304). In plants, the biochemical synthesis of 2-hydroxy-5-oxoproline has been known but has been poorly characterized. Moreover, the function of 2-hydroxy-5-oxoproline in plants and the significance of its pool size (tissue concentration) are unknown. Finally, the art provides no specific guidance as to precisely what transaminase(s) or hydrolase(s) may exist and/or be active in catalyzing the synthesis of 2-hydroxy-5-oxoproline in plants, and no such plant transaminases have been reported, isolated or characterized.

SUMMARY OF THE INVENTION

The invention relates to transgenic plants exhibiting enhanced growth rates, seed and fruit yields, and overall biomass yields, as well as methods for generating growth-enhanced transgenic plants. In one embodiment, transgenic plants engineered to over-express glutamine phenylpyruvate transaminase (GPT) are provided. In general, these plants out-grow their wild-type counterparts by about 50%.

Applicants have identified the enzyme glutamine phenylpyruvate transaminase (GPT) as a catalyst of 2-hydroxy-5-oxoproline (2-oxoglutaramate) synthesis in plants. 2-oxoglutaramate is a powerful signal metabolite which regulates the function of a large number of genes involved in the photosynthesis apparatus, carbon fixation and nitrogen metabolism.

By preferentially increasing the concentration of the signal metabolite 2-oxoglutaramate (i.e., in foliar tissues), the transgenic plants of the invention are capable of producing higher overall yields over shorter periods of time, and therefore may provide agricultural industries with enhanced productivity across a wide range of crops. Importantly, unlike many transgenic plants described to date, the invention utilizes natural plant genes encoding a natural plant enzyme. The enhanced growth characteristics of the transgenic plants of the invention are achieved essentially by introducing additional GPT capacity into the plant. Thus, the transgenic plants of the invention do not express any toxic substances, growth hormones, viral or bacterial gene products, and are therefore free of many of the concerns that have heretofore impeded the adoption of transgenic plants in certain parts of the World.

In one embodiment, the invention provides a transgenic plant comprising a GPT transgene, wherein said GPT transgene is operably linked to a plant promoter. In a specific embodiment, the GPT transgene encodes a polypeptide having an amino acid sequence selected from the group consisting of (a) SEQ ID NO: 2; SEQ ID NO: 9; SEQ ID NO: 15, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO 24, SEQ ID NO: 30, SEQ ID NO:31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35 and SEQ ID NO: 36, and (b) an amino acid sequence that is at least 75% identical to any one of SEQ ID NO: 2; SEQ ID NO: 9; SEQ ID NO: 15, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO 24, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35 and SEQ ID NO: 36 and has GPT activity.

In some embodiments, the GPT transgene is incorporated into the genome of the plant. The transgenic plant of the invention may be a monocotyledonous or a dicotyledonous plant.

The invention also provides progeny of any generation of the transgenic plants of the invention, wherein said progeny comprises a GPT transgene, as well as a seed of any generation of the transgenic plants of the invention, wherein said seed comprises said GPT transgene. The transgenic plants of the invention may display one or more enhanced growth characteristics when compared to an analogous wild-type or untransformed plant, including without limitation increased growth rate, increased biomass yield, increased seed yield, increased flower or flower bud yield, increased fruit or pod yield, larger leaves, and increased levels of GPT activity and/or increased levels of 2-oxoglutaramate. In some embodiments, the transgenic plants of the invention display increased nitrogen use efficiency.

Methods for producing the transgenic plants of the invention and seeds thereof are also provided, including methods for producing a plant having enhanced growth characteristics, increased nitrogen use efficiency and increased tolerance to germination or growth in salt or saline conditions, relative to an analogous wild type or untransformed plant.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1. Nitrogen assimilation and 2-oxoglutaramate biosynthesis: schematic of metabolic pathway.

FIG. 2. Photograph showing comparison of transgenic tobacco plants over-expressing GPT, compared to wild type tobacco plant. From left to right: wild type plant, Alfalfa GS1 transgene, *Arabidopsis* GPT transgene. See Example 3, infra.

FIG. 3. Photograph showing comparison of transgenic Micro-Tom tomato plants over-expressing GPT, compared to wild type tomato plant. (A) wild type plant; (B) *Arabidopsis* GPT transgene. See Example 4, infra.

FIG. 4. Photograph showing comparisons of leaf sizes between wild type (top leaf) and GPT transgenic (bottom leaf) tobacco plants.

DETAILED DESCRIPTION OF THE INVENTION Definitions

Unless otherwise defined, all terms of art, notations and other scientific terminology used herein are intended to have the meanings commonly understood by those of skill in the art to which this invention pertains. In some cases, terms with commonly understood meanings are defined herein for clarity and/or for ready reference, and the inclusion of such definitions herein should not necessarily be construed to represent a substantial difference over what is generally understood in the art. The techniques and procedures described or referenced herein are generally well understood and commonly employed using conventional methodology by those skilled in the art, such as, for example, the widely utilized molecular cloning methodologies described in Sambrook et al., Molecular Cloning: A Laboratory Manual 3rd. edition (2001) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; Current Protocols in Molecular Biology (Ausbel et al., eds., John Wiley & Sons, Inc. 2001; Transgenic Plants: Methods and Protocols (Leandro Pena, ed., Humana Press, 1st edition, 2004); and, *Agrobacterium* Protocols (Wan, ed., Humana Press, 2nd edition, 2006). As appropriate, procedures involving the use of

commercially available kits and reagents are generally carried out in accordance with manufacturer defined protocols and/or parameters unless otherwise noted.

The term “nucleic acid” refers to deoxyribonucleotides or ribonucleotides and polymers thereof (“polynucleotides”) in either single- or double-stranded form. Unless specifically limited, the term “polynucleotide” encompasses nucleic acids containing known analogues of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g. degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., 1991, Nucleic Acid Res. 19: 5081; Ohtsuka et al., 1985 J. Biol. Chem. 260: 2605-2608; and Cassol et al., 1992; Rossolini et al., 1994, Mol. Cell. Probes 8: 91-98). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

The term “promoter” refers to a nucleic acid control sequence or sequences that direct transcription of an operably linked nucleic acid. As used herein, a “plant promoter” is a promoter that functions in plants. Promoters include necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A “constitutive” promoter is a promoter that is active under most environmental and developmental conditions. An “inducible” promoter is a promoter that is active under environmental or developmental regulation. The term “operably linked” refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

The terms “polypeptide,” “peptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers.

The term “amino acid” refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an a carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid.

Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

The term “plant” includes whole plants, plant organs (e.g., leaves, stems, flowers, roots, reproductive organs, embryos and parts thereof, etc.), seedlings, seeds and plant cells and progeny thereof. The class of plants which can be used in the method of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous.

The terms “GPT polynucleotide” and “GPT nucleic acid” are used interchangeably herein, and refer to a full length or partial length polynucleotide sequence of a gene which encodes a polypeptide involved in catalyzing the synthesis of 2-oxoglutaramate, and includes polynucleotides containing both translated (coding) and un-translated sequences, as well as the complements thereof. The term “GPT coding sequence” refers to the part of the gene which is transcribed and encodes a GPT protein. The term “targeting sequence” refers to the amino terminal part of a protein which directs the protein into a subcellular compartment of a cell, such as a chloroplast in a plant cell. GPT polynucleotides are further defined by their ability to hybridize under defined conditions to the GPT polynucleotides specifically disclosed herein, or to PCR products derived therefrom.

A “GPT transgene” is a nucleic acid molecule comprising a GPT polynucleotide which is exogenous to transgenic plant, or plant embryo, organ or seed, harboring the nucleic acid molecule, or which is exogenous to an ancestor plant, or plant embryo, organ or seed thereof, of a transgenic plant harboring the GPT polynucleotide.

Exemplary GPT polynucleotides of the invention are presented herein, and include GPT coding sequences for *Arabidopsis*, Rice, Barley, Bamboo, Soybean, Grape, and Zebra Fish GPTs.

Partial length GPT polynucleotides include polynucleotide sequences encoding N- or C-terminal truncations of GPT, mature GPT (without targeting sequence) as well as sequences encoding domains of GPT. Exemplary GPT polynucleotides encoding N-terminal truncations of GPT include *Arabidopsis* −30, −45 and −56 constructs, in which coding sequences for the first 30, 45, and 56, respectively, amino acids of the full length GPT structure of SEQ ID NO: 2 are eliminated.

In employing the GPT polynucleotides of the invention in the generation of transformed cells and transgenic plants, one of skill will recognize that the inserted polynucleotide sequence need not be identical, but may be only “substantially identical” to a sequence of the gene from which it was derived, as further defined below. The term “GPT polynucleotide” specifically

encompasses such substantially identical variants. Similarly, one of skill will recognize that because of codon degeneracy, a number of polynucleotide sequences will encode the same polypeptide, and all such polynucleotide sequences are meant to be included in the term GPT polynucleotide. In addition, the term specifically includes those sequences substantially identical (determined as described below) with an GPT polynucleotide sequence disclosed herein and that encode polypeptides that are either mutants of wild type GPT polypeptides or retain the function of the GPT polypeptide (e.g., resulting from conservative substitutions of amino acids in a GPT polypeptide). The term "GPT polynucleotide" therefore also includes such substantially identical variants.

The term "conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The following eight groups each contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (see, e.g., Creighton, Proteins (1984)).

Macromolecular structures such as polypeptide structures can be described in terms of various levels of organization. For a general discussion of this organization, see, e.g., Alberts et al., *Molecular Biology of the Cell* (3rd ed., 1994) and Cantor and Schimmel, *Biophysical Chemistry Part I: The Conformation of Biological Macromolecules* (1980). "Primary structure" refers

to the amino acid sequence of a particular peptide. “Secondary structure” refers to locally ordered, three dimensional structures within a polypeptide. These structures are commonly known as domains. Domains are portions of a polypeptide that form a compact unit of the polypeptide and are typically 25 to approximately 500 amino acids long. Typical domains are made up of sections of lesser organization such as stretches of β-sheet and α-helices. “Tertiary structure” refers to the complete three dimensional structure of a polypeptide monomer. “Quaternary structure” refers to the three dimensional structure formed by the noncovalent association of independent tertiary units. Anisotropic terms are also known as energy terms.

The term “isolated” refers to material which is substantially or essentially free from components which normally accompany the material as it is found in its native or natural state. However, the term “isolated” is not intended refer to the components present in an electrophoretic gel or other separation medium. An isolated component is free from such separation media and in a form ready for use in another application or already in use in the new application/milieu. An “isolated” antibody is one that has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would interfere with diagnostic or therapeutic uses for the antibody, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. In preferred embodiments, the antibody will be purified (1) to greater than 95% by weight of antibody as determined by the Lowry method, and most preferably more than 99% by weight, (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator, or (3) to homogeneity by SDS-PAGE under reducing or nonreducing conditions using Coomassie blue or, preferably, silver stain. Isolated antibody includes the antibody in situ within recombinant cells since at least one component of the antibody's natural environment will not be present. Ordinarily, however, isolated antibody will be prepared by at least one purification step.

The term “heterologous” when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, a nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, e.g., a nucleic acid encoding a protein from one source and a nucleic acid encoding a peptide sequence from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion protein).

The terms “identical” or percent “identity,” in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., about 70% identity, preferably 75%, 80%, 85%, 90%, or 95% identity over a specified region, when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using a sequence comparison algorithms, or by manual alignment and visual inspection. This definition also refers to the complement of a test sequence, which has substantial sequence or subsequence complementarity when the test sequence has substantial identity to a reference sequence. This definition also refers to the complement of a test sequence, which has

substantial sequence or subsequence complementarity when the test sequence has substantial identity to a reference sequence.

When percentage of sequence identity is used in reference to polypeptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the polypeptide. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A “comparison window”, as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, 1981, Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman & Wunsch, 1970, J. Mol. Biol. 48:443, by the search for similarity method of Pearson & Lipman, 1988, Proc. Nat'l. Acad. Sci. USA 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by manual alignment and visual inspection (see, e.g., Current Protocols in Molecular Biology (Ausubel et al., eds. 1995 supplement)).

A preferred example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., 1977, Nuc. Acids Res. 25:3389-3402 and Altschul et al., 1990, J. Mol. Biol. 215:403-410, respectively. BLAST and BLAST 2.0 are used, typically with the default parameters described herein, to determine percent sequence identity for the nucleic acids and proteins of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores

are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, M=5, N=−4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word length of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=−4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, 1993, Proc. Nat'l. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

The phrase “stringent hybridization conditions” refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Probes, “Overview of principles of hybridization and the strategy of nucleic acid assays” (1993). Generally, highly stringent conditions are selected to be about 5-10° C. lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30° C. below the Tm. Tm is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at Tm, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0M sodium ion, typically about 0.01 to 1.0M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C. for short probes (e.g., 10 to 50 nucleotides) and at least about 60° C. for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the

maximum codon degeneracy permitted by the genetic code. In such cased, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

Genomic DNA or cDNA comprising GPT polynucleotides may be identified in standard Southern blots under stringent conditions using the GPT polynucleotide sequences disclosed here. For this purpose, suitable stringent conditions for such hybridizations are those which include a hybridization in a buffer of 40% formamide, 1M NaCl, 1% SDS at 37° C., and at least one wash in 0.2×SSC at a temperature of at least about 50° C., usually about 55° C. to about 60° C., for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions may be utilized to provide conditions of similar stringency.

A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, e.g., a northern or Southern blot.

Transgenic Plants:

The invention provides novel transgenic plants exhibiting substantially enhanced growth and other agronomic characteristics, including without limitation faster growth, greater mature plant fresh weight and total biomass, earlier and more abundant flowering, and greater fruit and seed yields. The transgenic plants of the invention are generated by introducing into a plant one or more expressible genetic constructs capable of driving the expression of one or more polynucleotides encoding glutamine phenylpyruvate transaminase (GPT). The invention is exemplified, for example, by the generation of transgenic tobacco plants carrying and expressing the heterologous *Arabidopsis* GPT gene (Example 2, infra). It is expected that all plant species also contain a GPT homolog which functions in the same metabolic pathway, namely the biosynthesis of the signal metabolite 2-hydroxy-5-oxoproline. Thus, in the practice of the invention, any plant gene encoding a GPT homolog or functional variants thereof may be useful in the generation of transgenic plants of this invention.

In stable transformation embodiments of the invention, one or more copies of the expressible genetic construct become integrated into the host plant genome, thereby providing increased GPT enzyme capacity into the plant, which serves to mediate increased synthesis of 2-oxoglutaramate, which in turn signals metabolic gene expression, resulting in increased plant growth and the enhancement of plant growth and other agronomic characteristics. 2-oxoglutaramate is a metabolite which is an extremely potent effector of gene expression, metabolism and plant growth (U.S. Pat. No. 6,555,500), and which may play a pivotal role in the coordination of the carbon and nitrogen metabolism systems (Lancien et al., 2000*, Enzyme Redundancy and the Importance of* 2*-Oxoglutarate in Higher Plants Ammonium Assimilation*, Plant Physiol. 123: 817-824). See, also, the schematic of the 2-oxoglutaramate pathway shown in FIG. 1.

In one aspect of the invention, applicants have isolated a nucleic acid molecule encoding the *Arabidopsis* glutamine phenylpyruvate transaminase (GPT) enzyme (see Example 1, infra), and have demonstrated for the first time that the

expressed recombinant enzyme is active and capable of catalyzing the synthesis of the signal metabolite, 2-oxoglutaramate (Example 2, infra). Further, applicants have demonstrated for the first time that over-expression of the *Arabidopsis* glutamine transaminase gene in a transformed heterologous plant results in enhanced CO_2 fixation rates and increased growth characteristics (Example 3, infra).

As disclosed herein (see Example 3, infra), over-expression of a transgene comprising the full-length *Arabidopsis* GPT coding sequence in transgenic tobacco plants also results in faster CO_2 fixation, and increased levels of total protein, glutamine and 2-oxoglutaramate. These transgenic plants also grow faster than wild-type plants (FIG. 2). Similarly, in studies conducted with tomato plants (see Example 4, infra), tomato plants transformed with the *Arabidopsis* GPT transgene showed significant enhancement of growth rate, flowering, and seed yield in relation to wild type control plants (FIG. 3 and Example 4, infra).

In addition to the transgenic tobacco plants referenced above, various other species of transgenic plants comprising GPT and showing enhanced growth characteristics have been generated in two species of Tomato, Pepper, Beans, Cowpea, Alfalfa, Cantaloupe, Pumpkin, *Arabidopsis* and Camilena (see co-owned, co-pending U.S. application Ser. No. 12/551,271, filed Aug. 31, 2009, the contents of which are incorporated herein by reference in its entirety). The foregoing transgenic plants were generated using a variety of transformation methodologies, including *Agrobacterium*-mediated callus, floral dip, seed inoculation, pod inoculation, and direct flower inoculation, as well as combinations thereof, and via sexual crosses of single transgene plants, using various GPT transgenes.

The invention also provides methods of generating a transgenic plant having enhanced growth and other agronomic characteristics. In one embodiment, a method of generating a transgenic plant having enhanced growth and other agronomic characteristics comprises introducing into a plant cell an expression cassette comprising a nucleic acid molecule encoding a GPT transgene, under the control of a suitable promoter capable of driving the expression of the transgene, so as to yield a transformed plant cell, and obtaining a transgenic plant which expresses the encoded GPT. In another embodiment, a method of generating a transgenic plant having enhanced growth and other agronomic characteristics comprises introducing into a plant cell one or more nucleic acid constructs or expression cassettes comprising nucleic acid molecules encoding a GPT transgene, under the control of one or more suitable promoters (and, optionally, other regulatory elements) capable of driving the expression of the transgenes, so as to yield a plant cell transformed thereby, and obtaining a transgenic plant which expresses the GPT transgene to produce a biologically active GPT protein.

Any number of GPT polynucleotides may be used to generate the transgenic plants of the invention. GPT proteins are highly conserved among various plant species, and it is evident from the experimental data disclosed herein that closely-related non-plant GPTs may be used as well (e.g., *Danio rerio* GPT). With respect to GPT, numerous GPT polynucleotides derived from different species have been shown to be active and useful as GPT transgenes.

In a specific embodiment, the GPT transgene is a GPT polynucleotide encoding an *Arabidopsis* derived GPT, such as the GPT of SEQ ID NO: 2, SEQ ID NO: 21 and SEQ ID NO: 30. The GPT transgene may be encoded by the nucleotide sequence of SEQ ID NO: 1; a nucleotide sequence having at least 75% and more preferably at least 80% identity to SEQ ID NO: 1, and encoding a polypeptide having GPT activity; a nucleotide sequence encoding the polypeptide of SEQ ID NO: 2, or a polypeptide having at least 75% and more preferably at least 80% sequence identity thereto which has GPT activity; and a nucleotide sequence encoding the polypeptide of SEQ ID NO: 2 truncated at its amino terminus by between 30 to 56 amino acid residues, or a polypeptide having at least 75% and more preferably at least 80% sequence identity thereto which has GPT activity.

In another specific embodiment, the GPT transgene is a GPT polynucleotide encoding a Grape derived GPT, such as the Grape GPTs of SEQ ID NO: 9 and SEQ ID NO: 31. The GPT transgene may be encoded by the nucleotide sequence of SEQ ID NO: 8; a nucleotide sequence having at least 75% and more preferably at least 80% identity to SEQ ID NO: 8, and encoding a polypeptide having GPT activity; a nucleotide sequence encoding the polypeptide of SEQ ID NO: 9 or SEQ ID NO: 31, or a polypeptide having at least 75% and more preferably at least 80% sequence identity thereto which has GPT activity.

In yet another specific embodiment, the GPT transgene is a GPT polynucleotide encoding a Rice derived GPT, such as the Rice GPTs of SEQ ID NO: 11 and SEQ ID NO: 32. The GPT transgene may be encoded by the nucleotide sequence of SEQ ID NO: 10; a nucleotide sequence having at least 75% and more preferably at least 80% identity to SEQ ID NO: 10, and encoding a polypeptide having GPT activity; a nucleotide sequence encoding the polypeptide of SEQ ID NO: 11 or SEQ ID NO: 32, or a polypeptide having at least 75% and more preferably at least 80% sequence identity thereto which has GPT activity.

In yet another specific embodiment, the GPT transgene is a GPT polynucleotide encoding a Soybean derived GPT, such as the Soybean GPTs of SEQ ID NO: 13, SEQ IS NO: 33 or SEQ ID NO: 33 with a further Isoleucine at the N-terminus of the sequence. The GPT transgene may be encoded by the nucleotide sequence of SEQ ID NO: 12; a nucleotide sequence having at least 75% and more preferably at least 80% identity to SEQ ID NO: 12, and encoding a polypeptide having GPT activity; a nucleotide sequence encoding the polypeptide of SEQ ID NO: 13 or SEQ ID NO: 33 or SEQ ID NO: 33 with a further Isoleucine at the N-terminus of the sequence, or a polypeptide having at least 75% and more preferably at least 80% sequence identity thereto which has GPT activity.

In yet another specific embodiment, the GPT transgene is a GPT polynucleotide encoding a Barley derived GPT, such as the Barley GPTs of SEQ ID NO: 15 and SEQ ID NO: 34. The GPT transgene may be encoded by the nucleotide sequence of SEQ ID NO: 14; a nucleotide sequence having at least 75% and more preferably at least 80% identity to SEQ ID NO: 10, and encoding a polypeptide having GPT activity; a nucleotide sequence encoding the polypeptide of SEQ ID NO: 15 or SEQ ID NO: 34, or a polypeptide having at least 75% and more preferably at least 80% sequence identity thereto which has GPT

activity.

In yet another specific embodiment, the GPT transgene is a GPT polynucleotide encoding a Zebra fish derived GPT, such as the Zebra fish GPTs of SEQ ID NO: 17 and SEQ ID NO: 35. The GPT transgene may be encoded by the nucleotide sequence of SEQ ID NO: 16; a nucleotide sequence having at least 75% and more preferably at least 80% identity to SEQ ID NO: 16, and encoding a polypeptide having GPT activity; a nucleotide sequence encoding the polypeptide of SEQ ID NO: 17 or SEQ ID NO: 35, or a polypeptide having at least 75% and more preferably at least 80% sequence identity thereto which has GPT activity.

In yet another specific embodiment, the GPT transgene is a GPT polynucleotide encoding a Bamboo derived GPT, such as the Bamboo GPT of SEQ ID NO: 36. The GPT transgene may be encoded by a nucleotide sequence encoding the polypeptide of SEQ ID NO: 36, or a polypeptide having at least 75% and more preferably at least 80% sequence identity thereto which has GPT activity.

As will be appreciated by one skilled in the art, other GPT polynucleotides suitable for use as GPT transgenes in the practice of the invention may be obtained by various means, and tested for the ability to direct the expression of a GPT with GPT activity in a recombinant expression system (i.e., *E. coli* (see Examples 20-23), in a transient in planta expression system (see Example 19), or in a transgenic plant (see Examples 1-18).

Transgene Constructs/Expression Vectors

In order to generate the transgenic plants of the invention, the gene coding sequence for the desired transgene(s) must be incorporated into a nucleic acid construct (also interchangeably referred to herein as a/an (transgene) expression vector, expression cassette, expression construct or expressible genetic construct), which can direct the expression of the transgene sequence in transformed plant cells. Such nucleic acid constructs carrying the transgene(s) of interest may be introduced into a plant cell or cells using a number of methods known in the art, including but not limited to electroporation, DNA bombardment or biolistic approaches, microinjection, and via the use of various DNA-based vectors such as *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes* vectors. Once introduced into the transformed plant cell, the nucleic acid construct may direct the expression of the incorporated transgene(s) (i.e., GPT), either in a transient or stable fashion. Stable expression is preferred, and is achieved by utilizing plant transformation vectors which are able to direct the chromosomal integration of the transgene construct. Once a plant cell has been successfully transformed, it may be cultivated to regenerate a transgenic plant.

A large number of expression vectors suitable for driving the constitutive or induced expression of inserted genes in transformed plants are known. In addition, various transient expression vectors and systems are known. To a large extent, appropriate expression vectors are selected for use in a particular method of gene transformation (see, infra). Broadly speaking, a typical plant expression vector for generating transgenic plants will comprise the transgene of interest under the

expression regulatory control of a promoter, a selectable marker for assisting in the selection of transformants, and a transcriptional terminator sequence.

More specifically, the basic elements of a nucleic acid construct for use in generating the transgenic plants of the invention are: a suitable promoter capable of directing the functional expression of the transgene(s) in a transformed plant cell, the transgene(s) (i.e., GPT coding sequence) operably linked to the promoter, preferably a suitable transcription termination sequence (i.e., nopaline synthetic enzyme gene terminator) operably linked to the transgene, and sometimes other elements useful for controlling the expression of the transgene, as well as one or more selectable marker genes suitable for selecting the desired transgenic product (i.e., antibiotic resistance genes).

As *Agrobacterium tumefaciens* is the primary transformation system used to generate transgenic plants, there are numerous vectors designed for *Agrobacterium* transformation. For stable transformation, *Agrobacterium* systems utilize “binary” vectors that permit plasmid manipulation in both *E. coli* and *Agrobacterium*, and typically contain one or more selectable markers to recover transformed plants (Hellens et al., 2000*, Technical focus: A guide to Agrobacterium binary Ti vectors*. Trends Plant Sci 5:446-451). Binary vectors for use in *Agrobacterium* transformation systems typically comprise the borders of T-DNA, multiple cloning sites, replication functions for *Escherichia coli* and *A. tumefaciens*, and selectable marker and reporter genes.

So-called “super-binary” vectors provide higher transformation efficiencies, and generally comprise additional virulence genes from a Ti (Komari et al., 2006, Methods Mol. Biol. 343: 15-41). Super binary vectors are typically used in plants which exhibit lower transformation efficiencies, such as cereals. Such additional virulence genes include without limitation virB, virE, and virG (Vain et al., 2004*, The effect of additional virulence genes on transformation efficiency, transgene integration and expression in rice plants using the pGreen/pSoup dual binary vector system*. Transgenic Res. 13: 593-603; Srivatanakul et al., 2000*, Additional virulence genes influence transgene expression: transgene copy number, integration pattern and expression*. J. Plant Physiol. 157, 685-690; Park et al., 2000*, Shorter T-DNA or additional virulence genes improve Agrobacterium-mediated transformation*. Theor. Appl. Genet. 101, 1015-1020; Jin et al., 1987*, Genes responsible for the supervirulence phenotype of Agrobacterium tumefaciens A*281. J. Bacteriol. 169: 4417-4425).

In the embodiments exemplified herein (see Examples, infra), expression vectors which place the inserted transgene(s) under the control of the constitutive CaMV 35S promoter are employed. A number of expression vectors which utilize the CaMV 35S promoter are known and/or commercially available. However, numerous promoters suitable for directing the expression of the transgene are known and may be used in the practice of the invention, as further described, infra.

Plant Promoters

A large number of promoters which are functional in plants are known in the art. In constructing GPT transgene constructs, the selected promoter(s) may be constitutive, non-specific promoters such as the Cauliflower Mosaic Virus 35S ribosomal

promoter (CaMV 35S promoter), which is widely employed for the expression of transgenes in plants. Examples of other strong constitutive promoters include without limitation the rice actin 1 promoter, the CaMV 19S promoter, the Ti plasmid nopaline synthase promoter, the alcohol dehydrogenase promoter and the sucrose synthase promoter.

Alternatively, in some embodiments, it may be desirable to select a promoter based upon the desired plant cells to be transformed by the transgene construct, the desired expression level of the transgene, the desired tissue or subcellular compartment for transgene expression, the developmental stage targeted, and the like.

For example, when expression in photosynthetic tissues and compartments is desired, a promoter of the ribulose bisphosphate carboxylase (RuBisCo) gene may be employed. When the expression in seeds is desired, promoters of various seed storage protein genes may be employed. For expression in fruits, a fruit-specific promoter such as tomato 2A11 may be used. Examples of other tissue specific promoters include the promoters encoding lectin (Vodkin et al., 1983, Cell 34:1023-31; Lindstrom et al., 1990, Developmental Genetics 11:160-167), corn alcohol dehydrogenase 1 (Vogel et al, 1989, J. Cell. Biochem. (Suppl. 0) 13:Part D; Dennis et al., 1984, Nucl. Acids Res., 12(9): 3983-4000), corn light harvesting complex (Simpson, 1986, Science, 233: 34-38; Bansal et al., 1992, Proc. Natl. Acad. Sci. USA, 89: 3654-3658), corn heat shock protein (Odell et al., 1985, Nature, 313: 810-812; Rochester et al., 1986, EMBO J., 5: 451-458), pea small subunit RuBP carboxylase (Poulsen et al., 1986, Mol. Gen. Genet., 205(2): 193-200; Cashmore et al., 1983, Gen. Eng. Plants, Plenum Press, New York, pp 29-38), Ti plasmid mannopine synthase and Ti plasmid nopaline synthase (Langridge et al., 1989, Proc. Natl. Acad. Sci. USA, 86: 3219-3223), petunia chalcone isomerase (Van Tunen et al., 1988, EMBO J. 7(5): 1257-1263), bean glycine rich protein 1 (Keller et al., 1989, EMBO J. 8(5): 1309-1314), truncated CaMV 35S (Odell et al., 1985, supra), potato patatin (Wenzler et al., 1989, Plant Mol. Biol. 12: 41-50), root cell (Conkling et al., 1990, Plant Physiol. 93: 1203-1211), maize zein (Reina et al., 1990, Nucl. Acids Res. 18(21): 6426; Kriz et al., 1987, Mol. Gen. Genet. 207(1): 90-98; Wandelt and Feix, 1989, Nuc. Acids Res. 17(6): 2354; Langridge and Feix, 1983, Cell 34: 1015-1022; Reina et al., 1990, Nucl. Acids Res. 18(21): 6426), globulin-1 (Belanger and Kriz, 1991, Genetics 129: 863-872), α-tubulin (Carpenter et al., 1992, Plant Cell 4(5): 557-571; Uribe et al., 1998, Plant Mol. Biol. 37(6): 1069-1078), cab (Sullivan, et al., 1989, Mol. Gen. Genet. 215(3): 431-440), PEPCase (Hudspeth and Grula, 1989, Plant Mol. Biol. 12: 579-589), R gene complex (Chandler et al., 1989, The Plant Cell 1: 1175-1183), chalcone synthase (Franken et al., 1991, EMBO J. 10(9): 2605-2612) and glutamine synthetase promoters (U.S. Pat. No. 5,391,725; Edwards et al., 1990, Proc. Natl. Acad. Sci. USA 87: 3459-3463; Brears et al., 1991, Plant J. 1(2): 235-244).

In addition to constitutive promoters, various inducible promoter sequences may be employed in cases where it is desirable to regulate transgene expression as the transgenic plant regenerates, matures, flowers, etc. Examples of such inducible promoters include promoters of heat shock genes, protection responding genes (i.e., phenylalanine ammonia lyase; see, for example Bevan et al., 1989, EMBO J. 8(7): 899-906), wound responding genes (i.e., cell wall protein genes), chemically inducible genes (i.e., nitrate reductase, chitinase) and dark inducible genes (i.e., asparagine synthetase; see, for example U.S. Pat. No. 5,256,558). Also, a number of plant nuclear genes are activated by light, including gene families encoding the

major chlorophyll a/b binding proteins (cab) as well as the small subunit of ribulose-1,5-bisphosphate carboxylase (rbcS) (see, for example, Tobin and Silverthorne, 1985, Annu. Rev. Plant Physiol. 36: 569-593; Dean et al., 1989, Annu. Rev. Plant Physiol. 40: 415-439.).

Other inducible promoters include ABA- and turgor-inducible promoters, the auxin-binding protein gene promoter (Schwob et al., 1993, Plant J. 4(3): 423-432), the UDP glucose flavonoid glycosyl-transferase gene promoter (Ralston et al., 1988, Genetics 119(1): 185-197); the MPI proteinase inhibitor promoter (Cordero et al., 1994, Plant J. 6(2): 141-150), the glyceraldehyde-3-phosphate dehydrogenase gene promoter (Kohler et al., 1995, Plant Mol. Biol. 29(6): 1293-1298; Quigley et al., 1989, J. Mol. Evol. 29(5): 412-421; Martinez et al., 1989, J. Mol. Biol. 208(4): 551-565) and light inducible plastid glutamine synthetase gene from pea (U.S. Pat. No. 5,391,725; Edwards et al., 1990, supra).

For a review of plant promoters used in plant transgenic plant technology, see Potenza et al., 2004, In Vitro Cell. Devel. Biol—Plant, 40(1): 1-22. For a review of synthetic plant promoter engineering, see, for example, Venter, M., 2007, Trends Plant Sci., 12(3): 118-124.

Glutamine Phenylpyruvate Transaminase (GPT) Transgene

The present invention discloses for the first time that plants contain a glutamine phenylpyruvate transaminase (GPT) enzyme which is directly functional in the synthesis of the signal metabolite 2-hydroxy-5-oxoproline. Until now, no plant transaminase with a defined function has been described. Applicants have isolated and tested GPT polynucleotide coding sequences derived from several plant and animal species, and have successfully incorporated the gene into heterologous transgenic host plants which exhibit markedly improved growth characteristics, including faster growth, higher foliar protein content, and faster CO_2 fixation rates.

It is expected that all plant species contain a GPT which functions in the same metabolic pathway, involving the biosynthesis of the signal metabolite 2-hydroxy-5-oxoproline. Thus, in the practice of the invention, any plant gene encoding a GPT homolog or functional variants thereof may be useful in the generation of transgenic plants of this invention. Moreover, given the structural similarity between various plant GPT protein structures and the putative (-and biologically active) GPT homolog from *Danio rerio* (Zebra fish) (see Example 22), other non-plant GPT homologs may be used in preparing GPT transgenes for use in generating the transgenic plants of the invention. When individually compared (by BLAST alignment) to the *Arabidopsis* mature protein sequence provided in SEQ ID NO: 30, the following sequence identities and homologies (BLAST “positives”, including similar amino acids) were obtained for the following mature GPT protein sequences:

[SEQ ID]	ORIGIN	% IDENTITY	% POSITIVE
[31]	Grape	84	93

[32]	Rice	83	91
[33]	Soybean	83	93
[34]	Barley	82	91
[35]	Zebra fish	83	92
[36]	Bamboo	81	90
	Corn	79	90
	Castor	84	93
	Poplar	85	93

Underscoring the conserved nature of the structure of the GPT protein across most plant species, the conservation seen within the above plant species extends to the non-human putative GPTs from Zebra fish and *Chlamydomonas*. In the case of Zebra fish, the extent of identity is very high (83% amino acid sequence identity with the mature *Arabidopsis* GPT of SEQ ID NO: 30, and 92% homologous taking similar amino acid residues into account). The Zebra fish mature GPT was confirmed by expressing it in *E. coli* and demonstrating biological activity (synthesis of 2-oxoglutaramate).

In order to determine whether putative GPT homologs would be suitable for generating the growth-enhanced transgenic plants of the invention, one may express the coding sequence thereof in *E. coli* or another suitable host and determine whether the 2-oxoglutaramate signal metabolite is synthesized at increased levels (see Examples 19-23). Where such an increase is demonstrated, the coding sequence may then be introduced into both homologous plant hosts and heterologous plant hosts, and growth characteristics evaluated. Any assay that is capable of detecting 2-oxoglutaramate with specificity may be used for this purpose, including without limitation the NMR and HPLC assays described in Example 2, infra. In addition, assays which measure GPT activity directly may be employed.

Any plant GPT with 2-oxoglutaramate synthesis activity may be used to transform plant cells in order to generate transgenic plants of the invention. There appears to be a high level of structural homology among plant species, which appears to extend beyond plants, as evidenced by the close homology between various plant GPT proteins and the putative Zebra fish GPT homolog. Therefore, various plant GPT genes may be used to generate growth-enhanced transgenic plants in a variety of heterologous plant species. In addition, GPT transgenes expressed in a homologous plant would be expected to result in the desired enhanced-growth characteristics as well (i.e., rice glutamine transaminase over-expressed in transgenic rice plants), although it is possible that regulation within a homologous cell may attenuate the expression of the transgene in some fashion that may not be operable in a heterologous cell.

Transcription Terminators:

In preferred embodiments, a 3′ transcription termination sequence is incorporated downstream of the transgene in order to

direct the termination of transcription and permit correct polyadenylation of the mRNA transcript. Suitable transcription terminators are those which are known to function in plants, including without limitation, the nopaline synthase (NOS) and octopine synthase (OCS) genes of *Agrobacterium tumefaciens*, the T7 transcript from the octopine synthase gene, the 3′ end of the protease inhibitor I or II genes from potato or tomato, the CaMV 35S terminator, the tml terminator and the pea rbcS E9 terminator. In addition, a gene's native transcription terminator may be used. In specific embodiments, described by way of the Examples, infra, the nopaline synthase transcription terminator is employed.

Selectable Markers:

Selectable markers are typically included in transgene expression vectors in order to provide a means for selecting transformants. While various types of markers are available, various negative selection markers are typically utilized, including those which confer resistance to a selection agent that inhibits or kills untransformed cells, such as genes which impart resistance to an antibiotic (such as kanamycin, gentamycin, anamycin, hygromycin and hygromycinB) or resistance to a herbicide (such as sulfonylurea, gulfosinate, phosphinothricin and glyphosate). Screenable markers include, for example, genes encoding β-glucuronidase (Jefferson, 1987, Plant Mol. Biol. Rep 5: 387-405), genes encoding luciferase (Ow et al., 1986, Science 234: 856-859) and various genes encoding proteins involved in the production or control of anthocyanin pigments (See, for example, U.S. Pat. No. 6,573,432). The *E. coli* glucuronidase gene (gus, gusA or uidA) has become a widely used selection marker in plant transgenics, largely because of the glucuronidase enzyme's stability, high sensitivity and ease of detection (e.g., fluorometric, spectrophotometric, various histochemical methods). Moreover, there is essentially no detectable glucuronidase in most higher plant species.

Transformation Methodologies and Systems:

Various methods for introducing the transgene expression vector constructs of the invention into a plant or plant cell are well known to those skilled in the art, and any capable of transforming the target plant or plant cell may be utilized.

Agrobacterium-mediated transformation is perhaps the most common method utilized in plant transgenics, and protocols for *Agrobacterium*-mediated transformation of a large number of plants are extensively described in the literature (see, for example, *Agrobacterium Protocols*, Wan, ed., Humana Press, 2nd edition, 2006). *Agrobacterium tumefaciens* is a Gram negative soil bacteria that causes tumors (Crown Gall disease) in a great many dicot species, via the insertion of a small segment of tumor-inducing DNA (“T-DNA”, ‘transfer DNA’) into the plant cell, which is incorporated at a semi-random location into the plant genome, and which eventually may become stably incorporated there. Directly repeated DNA sequences, called T-DNA borders, define the left and the right ends of the T-DNA. The T-DNA can be physically separated from the remainder of the Ti-plasmid, creating a ‘binary vector’ system.

Agrobacterium transformation may be used for stably transforming dicots, monocots, and cells thereof (Rogers et al., 1986, Methods Enzymol., 118: 627-641; Hernalsteen et al., 1984, EMBO J., 3: 3039-3041; Hoykass-Van Slogteren et al., 1984,

Nature, 311: 763-764; Grimsley et al., 1987, Nature 325: 167-1679; Boulton et al., 1989, Plant Mol. Biol. 12: 31-40; Gould et al., 1991, Plant Physiol. 95: 426-434). Various methods for introducing DNA into *Agrobacteria* are known, including electroporation, freeze/thaw methods, and triparental mating. The most efficient method of placing foreign DNA into *Agrobacterium* is via electroporation (Wise et al., 2006*, Three Methods for the Introduction of Foreign DNA into Agrobacterium*, Methods in Molecular Biology, vol. 343: *Agrobacterium* Protocols, 2/e, volume 1; Ed., Wang, Humana Press Inc., Totowa, N.J., pp. 43-53). In addition, given that a large percentage of T-DNAs do not integrate, *Agrobacterium*-mediated transformation may be used to obtain transient expression of a transgene via the transcriptional competency of unincorporated transgene construct molecules (Helens et al., 2005, Plant Methods 1:13).

A large number of *Agrobacterium* transformation vectors and methods have been described (Karimi et al., 2002, Trends Plant Sci. 7(5): 193-5), and many such vectors may be obtained commercially (for example, Invitrogen, Carlsbad, Calif.). In addition, a growing number of "open-source" *Agrobacterium* transformation vectors are available (for example, pCambia vectors; Cambia, Canberra, Australia). See, also, subsection herein on TRANSGENE CONSTRUCTS, supra. In a specific embodiment described further in the Examples, a pMON316-based vector was used in the leaf disc transformation system of Horsch et. al. (Horsch et al., 1995, Science 227:1229-1231) to generate growth enhanced transgenic tobacco and tomato plants.

Other commonly used transformation methods that may be employed in generating the transgenic plants of the invention include, without limitation, microprojectile bombardment, or biolistic transformation methods, protoplast transformation of naked DNA by calcium, polyethylene glycol (PEG) or electroporation (Paszkowski et al., 1984, EMBO J. 3: 2727-2722; Potrykus et al., 1985, Mol. Gen. Genet. 199: 169-177; Fromm et al., 1985, Proc. Nat. Acad. Sci. USA 82: 5824-5828; Shimamoto et al., 1989, Nature, 338: 274-276.

Biolistic transformation involves injecting millions of DNA-coated metal particles into target cells or tissues using a biolistic device (or "gene gun"), several kinds of which are available commercially. Once inside the cell, the DNA elutes off the particles and a portion may be stably incorporated into one or more of the cell's chromosomes (for review, see Kikkert et al., 2005*, Stable Transformation of Plant Cells by Particle Bombardment/Biolistics*, in: Methods in Molecular Biology, vol. 286: Transgenic Plants: Methods and Protocols, Ed. L. Peña, Humana Press Inc., Totowa, N.J.).

Electroporation is a technique that utilizes short, high-intensity electric fields to permeabilize reversibly the lipid bilayers of cell membranes (see, for example, Fisk and Dandekar, 2005*, Introduction and Expression of Transgenes in Plant Protoplasts*, in: Methods in Molecular Biology, vol. 286: Transgenic Plants: Methods and Protocols, Ed. L. Peña, Humana Press Inc., Totowa, N.J., pp. 79-90; Fromm et al., 1987*, Electroporation of DNA and RNA into plant protoplasts*, in Methods in Enzymology, Vol. 153, Wu and Grossman, eds., Academic Press, London, UK, pp. 351-366; Joersbo and Brunstedt, 1991*, Electroporation: mechanism and transient expression, stable transformation and biological effects in plant protoplasts*. Physiol. Plant. 81, 256-264; Bates, 1994*, Genetic transformation of plants by protoplast electroporation*. Mol. Biotech. 2:

135-145; Dillen et al., 1998*, Electroporation-mediated DNA transfer to plant protoplasts and intact plant tissues for transient gene expression assays*, in Cell Biology, Vol. 4, ed., Celis, Academic Press, London, UK, pp. 92-99). The technique operates by creating aqueous pores in the cell membrane, which are of sufficiently large size to allow DNA molecules (and other macromolecules) to enter the cell, where the transgene expression construct (as T-DNA) may be stably incorporated into plant genomic DNA, leading to the generation of transformed cells that can subsequently be regenerated into transgenic plants.

Newer transformation methods include so-called “floral dip” methods, which offer the promise of simplicity, without requiring plant tissue culture, as is the case with all other commonly used transformation methodologies (Bent et al., 2006, *Arabidopsis thaliana Floral Dip Transformation Method*, Methods Mol Biol, vol. 343: *Agrobacterium Protocols,* 2/e, volume 1; Ed., Wang, Humana Press Inc., Totowa, N.J., pp. 87-103; Clough and Bent, 1998*, Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana*, Plant J. 16: 735-743). However, with the exception of *Arabidopsis*, these methods have not been widely used across a broad spectrum of different plant species. Briefly, floral dip transformation is accomplished by dipping or spraying flowering plants in with an appropriate strain of *Agrobacterium tumefaciens*. Seeds collected from these T_0 plants are then germinated under selection to identify transgenic T_1 individuals. Example 16 demonstrated floral dip inoculation of *Arabidopsis* to generate transgenic *Arabidopsis* plants.

Other transformation methods include those in which the developing seeds or seedlings of plants are transformed using vectors such as *Agrobacterium* vectors. For example, such vectors may be used to transform developing seeds by injecting a suspension or mixture of the vector (i.e., *Agrobacteria*) directly into the seed cavity of developing pods (Wang and Waterhouse, 1997, Plant Mol. Biol. Reporter 15: 209-215). Seedlings may be transformed as described in Yasseem, 2009, Plant Mol. Biol. Reporter 27: 20-28. Germinating seeds may be transformed as described in Chee et al., 1989, Plant Pysiol. 91: 1212-1218. Intra-fruit methods, in which the vector is injected into fruit or developing fruit, may be also be used. Still other transformation methods include those in which the flower structure is targeted for vector inoculation, such as the flower inoculation methods.

The foregoing plant transformation methodologies may be used to introduce transgenes into a number of different plant cells and tissues, including without limitation, whole plants, tissue and organ explants including chloroplasts, flowering tissues and cells, protoplasts, meristem cells, callus, immature embryos and gametic cells such as microspores, pollen, sperm and egg cells, tissue cultured cells of any of the foregoing, any other cells from which a fertile regenerated transgenic plant may be generated. Callus is initiated from tissue sources including, but not limited to, immature embryos, seedling apical meristems, microspores and the like. Cells capable of proliferating as callus are also recipient cells for genetic transformation.

Methods of regenerating individual plants from transformed plant cells, tissues or organs are known and are described for numerous plant species.

As an illustration, transformed plantlets (derived from transformed cells or tissues) are cultured in a root-permissive growth medium supplemented with the selective agent used in the transformation strategy (i.e., and antibiotic such as kanamycin). Once rooted, transformed plantlets are then transferred to soil and allowed to grow to maturity. Upon flowering, the mature plants are preferably selfed (self-fertilized), and the resultant seeds harvested and used to grow subsequent generations. Examples 3-6 describe the regeneration of transgenic tobacco and tomato plants.

T_0 transgenic plants may be used to generate subsequent generations (e.g., T_1, T_2, etc.) by selfing of primary or secondary transformants, or by sexual crossing of primary or secondary transformants with other plants (transformed or untransformed).

Selection of Growth-Enhanced Transgenic Plants:

Transgenic plants may be selected, screened and characterized using standard methodologies. The preferred transgenic plants of the invention will exhibit one or more phenotypic characteristics indicative of enhanced growth and/or other desirable agronomic properties. Transgenic plants are typically regenerated under selective pressure in order to select transformants prior to creating subsequent transgenic plant generations. In addition, the selective pressure used may be employed beyond T_0 generations in order to ensure the presence of the desired transgene expression construct or cassette.

T_0 transformed plant cells, calli, tissues or plants may be identified and isolated by selecting or screening for the genetic composition of and/or the phenotypic characteristics encoded by marker genes contained in the transgene expression construct used for the transformation. For example, selection may be conducted by growing potentially-transformed plants, tissues or cells in a growth medium containing a growth-repressive amount of antibiotic or herbicide to which the transforming genetic construct can impart resistance. Further, the transformed plant cells, tissues and plants can be identified by screening for the activity of marker genes (i.e., β-glucuronidase) which may be present in the transgene expression construct.

Various physical and biochemical methods may be employed for identifying plants containing the desired transgene expression construct, as is well known. Examples of such methods include Southern blot analysis or various nucleic acid amplification methods (i.e., PCR) for identifying the transgene, transgene expression construct or elements thereof, Northern blotting, Si RNase protection, reverse transcriptase PCR (RT-PCR) amplification for detecting and determining the RNA transcription products, and protein gel electrophoresis, Western blotting, immunoprecipitation, enzyme immunoassay, and the like may be used for identifying the protein encoded and expressed by the transgene.

In another approach, expression levels of genes, proteins and/or metabolic compounds that are know to be modulated by transgene expression in the target plant may be used to identify transformants. In one embodiment of the present invention, increased levels of the signal metabolite 2-oxoglutaramate may be used to screen for desirable transformants.

Ultimately, the transformed plants of the invention may be screened for enhanced growth and/or other desirable agronomic characteristics. Indeed, some degree of phenotypic screening is generally desirable in order to identify transformed lines with the fastest growth rates, the highest seed yields, etc., particularly when identifying plants for subsequent selfing, cross-breeding and back-crossing. Various parameters may be used for this purpose, including without limitation, growth rates, total fresh weights, dry weights, seed and fruit yields (number, weight), seed and/or seed pod sizes, seed pod yields (e.g., number, weight), leaf sizes, plant sizes, increased flowering, time to flowering, overall protein content (in seeds, fruits, plant tissues), specific protein content (i.e., GS), nitrogen content, free amino acid, and specific metabolic compound levels (i.e., 2-oxoglutaramate). Generally, these phenotypic measurements are compared with those obtained from a parental identical or analogous plant line, an untransformed identical or analogous plant, or an identical or analogous wild-type plant (i.e., a normal or parental plant). Preferably, and at least initially, the measurement of the chosen phenotypic characteristic(s) in the target transgenic plant is done in parallel with measurement of the same characteristic(s) in a normal or parental plant. Typically, multiple plants are used to establish the phenotypic desirability and/or superiority of the transgenic plant in respect of any particular phenotypic characteristic.

Preferably, initial transformants are selected and then used to generate T_1 and subsequent generations by selfing (self-fertilization), until the transgene genotype breeds true (i.e., the plant is homozygous for the transgene). In practice, this is accomplished by screening at each generation for the desired traits and selfing those individuals, often repeatedly (i.e., 3 or 4 generations).

Stable transgenic lines may be crossed and back-crossed to create varieties with any number of desired traits, including those with stacked transgenes, multiple copies of a transgene, etc. Various common breeding methods are well know to those skilled in the art (see, e.g., Breeding Methods for Cultivar Development, Wilcox J. ed., American Society of Agronomy, Madison Wis. (1987)). Additionally, stable transgenic plants may be further modified genetically, by transforming such plants with further transgenes or additional copies of the parental transgene. Also contemplated are transgenic plants created by single transformation events which introduce multiple copies of a given transgene or multiple transgenes.

EXAMPLES

Various aspects of the invention are further described and illustrated by way of the several examples which follow, none of which are intended to limit the scope of the invention.

Example 1 Isolation of *Arabidopsis* Glutamine Phenylpyruvate Transaminase (GPT) Gene

In an attempt to locate a plant enzyme that is directly involved in the synthesis of the signal metabolite 2-oxoglutaramate, applicants hypothesized that the putative plant enzyme might bear some degree of structural relationship to a human protein that had been characterized as being involved in the synthesis of 2-oxoglutaramate. The human protein, glutamine transaminase K (E.C. 2.6.1.64) (also referred in the literature as cysteine conjugate 1-lyase, kyneurenine aminotransferase,

glutamine phenylpyruvate transaminase, and other names), had been shown to be involved in processing of cysteine conjugates of halogenated xenobiotics (Perry et al., 1995, FEBS Letters 360:277-280). Rather than having an activity involved in nitrogen assimilation, however, human cysteine conjugate β-lyase has a detoxifying activity in humans, and in animals (Cooper and Meister, 1977, supra). Nevertheless, the potential involvement of this protein in the synthesis of 2-oxoglutaramate was of interest.

Using the protein sequence of human cysteine conjugate β-lyase, a search against the TIGR *Arabidopsis* plant database of protein sequences identified one potentially related sequence, a polypeptide encoded by a partial sequence at the *Arabidopsis* gene locus at At1q77670, sharing approximately 36% sequence homology/identity across aligned regions.

The full coding region of the gene was then amplified from an *Arabidopsis* cDNA library (Stratagene) with the following primer pair:

SEQ ID NO: 321

5′-CCCATCGATGTACC TGGACATAAATGGTGTGATG-3′

SEQ ID NO: 322

5′-GATGGTACCTCAGACTTTTCTCTTAAGCTTCTGCTTC-3′

These primers were designed to incorporate Cla I (ATCGAT and Kpn I (GGTACC) restriction sites to facilitate subsequent subcloning into expression vectors for generating transgenic plants. Takara ExTaq DNA polymerase enzyme was used for high fidelity PCR using the following conditions: initial denaturing 94 C for 4 minutes, 30 cycles of 94° C. 30 second, annealing at 55° C. for 30 seconds, extension at 72° C. for 90 seconds, with a final extension of 72° C. for 7 minutes. The amplification product was digested with Cla I and Kpn I restriction enzymes, isolated from an agarose gel electrophoresis and ligated into vector pMon316 (Rogers, et. al. 1987 Methods in Enzymology 153:253-277) which contains the cauliflower mosaic virus (CaMV) 35S constitutive promoter and the nopaline synthase (NOS) 3′ terminator. The ligation product was transformed into DH5α cells and transformants sequenced to verify the insert.

A 1.3 kb cDNA was isolated and sequenced, and found to encode a full length protein of 440 amino acids in length, including a putative chloroplast signal sequence.

Example 2 Production of Biologically Active *Arabidopsis* Glutamine Phenylpyruvate Transaminase

To test whether the protein encoded by the cDNA isolated as described in Example 1, supra, is capable of catalyzing the synthesis of 2-oxoglutaramate, the cDNA was expressed in *E. coli*, purified, and assayed for its ability to synthesize 2-oxoglutaramate using a standard method.

NMR Assay for 2-Oxoglutaramate

Briefly, the resulting purified protein was added to a reaction mixture containing 150 mM Tris-HCl, pH 8.5, 1 mM beta mercaptoethanol, 200 mM glutamine, 100 mM glyoxylate and 200 μM pyridoxal 5′-phosphate. The reaction mixture without added test protein was used as a control. Test and control reaction mixtures were incubated at 37° C. for 20 hours, and then clarified by centrifugation to remove precipitated material. Supernatants were tested for the presence and amount of 2-oxoglutaramate using ^{13}C NMR with authentic chemically synthesized 2-oxoglutaramate as a reference. The products of the reaction are 2-oxoglutaramate and glycine, while the substrates (glutamine and glyoxylate) diminish in abundance. The cyclic 2-oxoglutaramate gives rise to a distinctive signal allowing it to be readily distinguished from the open chain glutamine precursor.

HPLC Assay for 2-Oxoglutaramate

An alternative assay for GPT activity uses HPLC to determine 2-oxoglutaramate production, following a modification of Calderon et al., 1985, J Bacteriol 161(2): 807-809. Briefly, a modified extraction buffer consisting of 25 mM Tris-HCl pH 8.5, 1 mM EDTA, 20 μM FAD, 10 mM Cysteine, and ˜1.5% (v/v) Mercaptoethanol. Tissue samples from the test material (i.e., plant tissue) are added to the extraction buffer at approximately a 1/3 ratio (w/v), incubated for 30 minutes at 37° C., and stopped with 200 μl of 20% TCA. After about 5 minutes, the assay mixture is centrifuged and the supernatant used to quantify 2-oxoglutaramate by HPLC, using an ION-300 7.8 mm ID×30 cm L column, with a mobile phase in 0.01N H_2SO_4, a flow rate of approximately 0.2 ml/min, at 40° C. Injection volume is approximately 20 μl, and retention time between about 38 and 39 minutes. Detection is achieved with 210 nm UV light.

Results Using NMR Assay:

This experiment revealed that the test protein was able to catalyze the synthesis of 2-oxoglutaramate. Therefore, these data indicate that the isolated cDNA encodes a glutamine phenylpyruvate transaminase that is directly involved in the synthesis of 2-oxoglutaramate in plants. Accordingly, the test protein was designated *Arabidopsis* glutamine phenylpyruvate transaminase, or “GPT”.

The nucleotide sequence of the *Arabidopsis* GPT coding sequence is shown in the Table of Sequences, SEQ ID NO. 1. The translated amino acid sequence of the GPT protein is shown in SEQ ID NO. 2.

Example 3 Creation of Transgenic Tobacco Plants Over-Expressing *Arabidopsis* GPT

Generation of Plant Expression Vector pMON-PJU:

Briefly, the plant expression vector pMon316-PJU was constructed as follows. The isolated cDNA encoding *Arabidopsis* GPT (Example 1) was cloned into the ClaI-KpnI polylinker site of the pMON316 vector, which places the GPT gene under the control of the constitutive cauliflower mosaic virus (CaMV) 35S promoter and the nopaline synthase (NOS)

transcriptional terminator. A kanamycin resistance gene was included to provide a selectable marker.

Agrobacterium-Mediated Plant Transformations:

pMON-PJU and a control vector pMon316 (without inserted DNA) were transferred to *Agrobacterium tumefaciens* strain pTiTT37ASE using a standard electroporation method (McCormac et al., 1998, Molecular Biotechnology 9:155-159), followed by plating on LB plates containing the antibiotics spectinomycin (100 micro gm/ml) and kanamycin (50 micro gm/ml). Antibiotic resistant colonies of *Agrobacterium* were examined by PCR to assure that they contained plasmid.

Nicotiana tabacum cv. Xanthi plants were transformed with pMON-PJU transformed *Agrobacteria* using the leaf disc transformation system of Horsch et. al. (Horsch et al., 1995, Science 227:1229-1231). Briefly, sterile leaf disks were inoculated and cultured for 2 days, then transferred to selective MS media containing 100 μg/ml kanamycin and 500 μg/ml clafaran. Transformants were confirmed by their ability to form roots in the selective media.

Generation of GPT Transgenic Tobacco Plants:

Sterile leaf segments were allowed to develop callus on Murashige & Skoog (M&S) media from which the transformant plantlets emerged. These plantlets were then transferred to the rooting-permissive selection medium (M&S medium with kanamycin as the selection agent). The healthy, and now rooted, transformed tobacco plantlets were then transferred to soil and allowed to grow to maturity and upon flowering the plants were selfed and the resultant seeds were harvested. During the growth stage the plants had been examined for growth phenotype and the CO_2 fixation rate was measured for many of the young transgenic plants.

Production of T1 and T2 Generation GPT Transgenic Plants:

Seeds harvested form the T_0 generation of the transgenic tobacco plants were germinated on M&S media containing kanamycin (100 mg/L) to enrich for the transgene. At least one fourth of the seeds did not germinate on this media (kanamycin is expected to inhibit germination of the seeds without resistance that would have been produced as a result of normal genetic segregation of the gene) and more than half of the remaining seeds were removed because of demonstrated sensitivity (even mild) to the kanamycin.

The surviving plants (T_1 generation) were thriving and these plants were then selfed to produce seeds for the T_2 generation. Seeds from the T_1 generation were germinated on MS media supplemented for the transformant lines with kanamycin (10 mg/liter). After 14 days they were transferred to sand and provided quarter strength Hoagland's nutrient solution supplemented with 25 mM potassium nitrate. They were allowed to grow at 24° C. with a photoperiod of 16 h light and 8 hr dark with a light intensity of 900 micromoles per meter squared per second. They were harvested 14 days after being transferred to the sand culture.

Characterization of GPT Transgenic Plants:

Harvested transgenic plants (both GPT transgenes and vector control transgenes) were analyzed for glutamine sythetase activity in root and leaf, whole plant fresh weight, total protein in root and leaf, and CO_2 fixation rate (Knight et al., 1988, Plant Physiol. 88: 333). Non-transformed, wild-type *A. tumefaciens* plants were also analyzed across the same parameters in order to establish a baseline control.

Growth characteristic results are tabulated below in Table I. Additionally, a photograph of the GPT transgenic plant compared to a wild type control plant is shown in FIG. 2 (together with GS1 transgenic tobacco plant). Across all parameters evaluated, the GPT transgenic tobacco plants showed enhanced growth characteristics. In particular, the GPT transgenic plants exhibited a greater than 50% increase in the rate of CO_2 fixation, and a greater than two-fold increase in glutamine synthetase activity in leaf tissue, relative to wild type control plants. In addition, the leaf-to-root GS ratio increased by almost three-fold in the transaminase transgenic plants relative to wild type control. Fresh weight and total protein quantity also increased in the transgenic plants, by about 50% and 80% (leaf), respectively, relative to the wild type control. These data demonstrate that tobacco plants overexpressing the *Arabidopsis* GPT transgene achieve significantly enhanced growth and CO_2 fixation rates.

TABLE I

Protein mg/gram fresh weight	Leaf	Root
Wild type - control	8.3	2.3
Line PN1-8 a second control	8.9	2.98
Line PN9-9	13.7	3.2
Glutamine Synthetase activity, micromoles/min/mg protein		
Wild type (Ratio of leaf:root = 4.1:1)	4.3	1.1
PN1-8 (Ratio of leaf:root = 4.2:1)	5.2	1.3
PN9-9 (Ratio of leaf:root = 10.9:1)	10.5	0.97
Whole Plant Fresh Weight, g		
Wild type	21.7	
PN1-8	26.1	
PN9-9	33.1	
CO_2 Fixation Rate,		

umole/m2/sec

Wild type	8.4
PN1-8	8.9
PN9-9	12.9

Data = average of three plants

Wild type - Control plants; not regenerated or transformed.

PN1 lines were produced by regeneration after transformation using a construct without inserted gene. A control against the processes of regeneration and transformation.

PN 9 lines were produced by regeneration after transformation using a construct with the *Arabidopsis* GPT gene.

Example 4 Generation of Transgenic Tomato Plants Carrying *Arabidopsis* GPT Transgene

Transgenic *Lycopersicon esculentum* (Micro-Tom Tomato) plants carrying the *Arabidopsis* GPT transgene were generated using the vectors and methods described in Example 3. T_0 transgenic tomato plants were generated and grown to maturity.

Initial growth characteristic data of the GPT transgenic tomato plants is presented in Table II. The transgenic plants showed significant enhancement of growth rate, flowering, and seed yield in relation to wild type control plants. In addition, the transgenic plants developed multiple main stems, whereas wild type plants developed with a single main stem. A photograph of a GPT transgenic tomato plant compared to a wild type plant is presented in FIG. 3.

TABLE II

Growth Characteristics	Wildtype Tomato	GPT Transgenic Tomato
Stem height, cm	6.5	18, 12, 11 major stems
Stems	1	3 major, 0 other
Buds	2	16
Flowers	8	12
Fruit	0	3

Example 5 Activity of Barley GPT Transgene in Planta

In this example, the putative coding sequence for Barley GPT was isolated and expressed from a transgene construct using

an in planta transient expression assay. Biologically active recombinant Barley GPT was produced, and catalyzed the increased synthesis of 2-oxoglutaramate, as confirmed by HPLC.

The Barley (*Hordeum vulgare*) GPT coding sequence was determined and synthesized. The DNA sequence of the Barley GPT coding sequence used in this example is provided in SEQ ID NO: 14, and the encoded GPT protein amino acid sequence is presented in SEQ ID NO: 15.

The coding sequence for Barley GPT was inserted into the 1305.1 cambia vector, and transferred to *Agrobacterium tumefaciens* strain LBA404 using a standard electroporation method (McCormac et al., 1998, Molecular Biotechnology 9:155-159), followed by plating on LB plates containing hygromycin (50 micro gm/ml). Antibiotic resistant colonies of *Agrobacterium* were selected for analysis.

The transient tobacco leaf expression assay consisted of injecting a suspension of transformed *Agrobacterium* (1.5-2.0 OD 650) into rapidly growing tobacco leaves. Intradermal injections were made in a grid across the leaf surface to assure that a significant amount of the leaf surface would be exposed to the *Agrobacterium*. The plant was then allowed to grow for 3-5 days when the tissue was extracted as described for all other tissue extractions and the GPT activity measured.

GPT activity in the inoculated leaf tissue (1217 nanomoles/gFWt/h) was three-fold the level measured in the control plant leaf tissue (407 nanomoles/gFWt/h), indicating that the *Hordeum* GPT construct directed the expression of biologically active GPT in a transgenic plant.

Example 6 Isolation and Expression of Recombinant Rice GPT Gene Coding Sequence and Analysis of Biological Activity

In this example, the putative coding sequence for rice GPT was isolated and expressed in *E. coli*. Biologically active recombinant rice GPT was produced, and catalyzed the increased synthesis of 2-oxoglutaramate, as confirmed by HPLC.

Materials and Methods:

Rice GPT Coding Sequence and Expression in *E. coli:*

The rice (*Oryza sativa*) GPT coding sequence was determined and synthesized, inserted into a PET28 vector, and expressed in *E. coli*. Briefly, *E. coli* cells were transformed with the expression vector and transformants grown overnight in LB broth diluted and grown to OD 0.4, expression induced with isopropyl-B-D-thiogalactoside (0.4 micromolar), grown for 3 hr and harvested. A total of 25×10^6 cells were then assayed for biological activity using the NMR assay, below. Untransformed, wild type *E. coli* cells were assayed as a control. An additional control used *E. coli* cells transformed with an empty vector.

The DNA sequence of the rice GPT coding sequence used in this example is provided in SEQ ID NO: 10, and the encoded

GPT protein amino acid sequence is presented in SEQ ID NO: 11.

HPLC Assay for 2-Oxoglutaramate:

HPLC was used to determine 2-oxoglutaramate production in GPT-overexpressing *E. coli* cells, following a modification of Calderon et al., 1985, J Bacteriol 161(2): 807-809. Briefly, a modified extraction buffer consisting of 25 mM Tris-HCl pH 8.5, 1 mM EDTA, 20 μM Pyridoxal phosphate, 10 mM Cysteine, and ˜1.5% (v/v) Mercaptoethanol was used. Samples (lysate from *E. coli* cells, 25×10^6 cells) were added to the extraction buffer at approximately a 1/3 ratio (w/v), incubated for 30 minutes at 37° C., and stopped with 200 μl of 20% TCA. After about 5 minutes, the assay mixture is centrifuged and the supernatant used to quantify 2-oxoglutaramate by HPLC, using an ION-300 7.8 mm ID×30 cm L column, with a mobile phase in 0.01 N H_2SO_4, a flow rate of approximately 0.2 ml/min, at 40° C. Injection volume is approximately 20 and retention time between about 38 and 39 minutes. Detection is achieved with 210 nm UV light.

NMR analysis comparison with authentic 2-oxoglutaramate was used to establish that the *Arabidopsis full length sequence expresses a GPT with* 2-oxoglutaramate synthesis activity. Briefly, authentic 2-oxoglutamate (structure confirmed with NMR) made by chemical synthesis to validate the HPLC assay, above, by confirming that the product of the assay (molecule synthesized in response to the expressed GPT) and the authentic 2-oxoglutaramate elute at the same retention time. In addition, when mixed together the assay product and the authentic compound elute as a single peak. Furthermore, the validation of the HPLC assay also included monitoring the disappearance of the substrate glutamine and showing that there was a 1:1 molar stoichiometry between glutamine consumed to 2-oxoglutamate produced. The assay procedure always included two controls, one without the enzyme added and one without the glutamine added. The first shows that the production of the 2-oxoglutaramate was dependent upon having the enzyme present, and the second shows that the production of the 2-oxoglutaramate was dependent upon the substrate glutamine.

Results:

Expression of the rice GPT coding sequence of SEQ ID NO: 10 resulted in the over-expression of recombinant GPT protein having 2-oxoglutaramate synthesis-catalyzing bioactivity. Specifically, 1.72 nanomoles of 2-oxoglutaramate activity was observed in the *E. coli* cells overexpressing the recombinant rice GPT, compared to only 0.02 nanomoles of 2-oxoglutaramate activity in control *E. coli* cells, an 86-fold activity level increase over control.

Example 7 Isolation and Expression of Recombinant Soybean GPT Gene Coding Sequence and Analysis of Biological Activity

In this example, the putative coding sequence for soybean GPT was isolated and expressed in *E. coli*. Biologically active recombinant soybean GPT was produced, and catalyzed the increased synthesis of 2-oxoglutaramate, as confirmed by HPLC.

Materials and Methods:

Soybean GPT Coding Sequence and Expression in *E. coli:*

The soybean (*Glycine max*) GPT coding sequence was determined and synthesized, inserted into a PET28 vector, and expressed in *E. coli*. Briefly, *E. coli* cells were transformed with the expression vector and transformants grown overnight in LB broth diluted and grown to OD 0.4, expression induced with isopropyl-B-D-thiogalactoside (0.4 micromolar), grown for 3 hr and harvested. A total of 25×10^6 cells were then assayed for biological activity using the HPLC assay, below. Untransformed, wild type *E. coli* cells were assayed as a control. An additional control used *E. coli* cells transformed with an empty vector.

The DNA sequence of the soybean GPT coding sequence used in this example is provided in SEQ ID NO: 12, and the encoded GPT protein amino acid sequence is presented in SEQ ID NO: 13.

HPLC Assay for 2-Oxoglutaramate:

HPLC was used to determine 2-oxoglutaramate production in GPT-overexpressing *E. coli* cells, as described in Example 6, supra.

Results:

Expression of the soybean GPT coding sequence of SEQ ID NO: 12 resulted in the over-expression of recombinant GPT protein having 2-oxoglutaramate synthesis-catalyzing bioactivity. Specifically, 31.9 nanomoles of 2-oxoglutaramate activity was observed in the *E. coli* cells overexpressing the recombinant soybean GPT, compared to only 0.02 nanomoles of 2-oxoglutaramate activity in control *E. coli* cells, a nearly 1.600-fold activity level increase over control.

Example 8 Isolation and Expression of Recombinant Zebra Fish GPT Gene Coding Sequence and Analysis of Biological Activity

In this example, the putative coding sequence for Zebra fish GPT was isolated and expressed in *E. coli*. Biologically active recombinant Zebra fish GPT was produced, and catalyzed the increased synthesis of 2-oxoglutaramate, as confirmed by HPLC.

Materials and Methods:

Zebra Fish GPT Coding Sequence and Expression in *E. coli:*

The Zebra fish (*Danio rerio*) GPT coding sequence was determined and synthesized, inserted into a PET28 vector, and expressed in *E. coli*. Briefly, *E. coli* cells were transformed with the expression vector and transformants grown overnight in

LB broth diluted and grown to OD 0.4, expression induced with isopropyl-B-D-thiogalactoside (0.4 micromolar), grown for 3 hr and harvested. A total of 25×10^6 cells were then assayed for biological activity using the HPLC assay, below. Untransformed, wild type *E. coli* cells were assayed as a control. An additional control used *E coli* cells transformed with an empty vector.

The DNA sequence of the Zebra fish GPT coding sequence used in this example is provided in SEQ ID NO: 16, and the encoded GPT protein amino acid sequence is presented in SEQ ID NO: 17.

HPLC Assay for 2-Oxoglutaramate:

HPLC was used to determine 2-oxoglutaramate production in GPT-overexpressing *E. coli* cells, as described in Example 6, supra.

Results:

Expression of the Zebra fish GPT coding sequence of SEQ ID NO: 16 resulted in the over-expression of recombinant GPT protein having 2-oxoglutaramate synthesis-catalyzing bioactivity. Specifically, 28.6 nanomoles of 2-oxoglutaramate activity was observed in the *E. coli* cells overexpressing the recombinant Zebra fish GPT, compared to only 0.02 nanomoles of 2-oxoglutaramate activity in control *E. coli* cells, a more than 1.400-fold activity level increase over control.

Example 9 Generation and Expression of Recombinant Truncated *Arabidopsis* GPT Gene Coding Sequences and Analysis of Biological Activity

In this example, two different truncations of the *Arabidopsis* GPT coding sequence were designed and expressed in *E. coli*, in order to evaluate the activity of GPT proteins in which the putative chloroplast signal peptide is absent or truncated. Recombinant truncated GPT proteins corresponding to the full length *Arabidopsis* GPT amino acid sequence of SEQ ID NO: 1, truncated to delete either the first 30 amino-terminal amino acid residues, or the first 45 amino-terminal amino acid residues, were successfully expressed and showed biological activity in catalyzing the increased synthesis of 2-oxoglutaramate, as confirmed by HPLC.

Materials and Methods:

Truncated *Arabidopsis* GPT Coding Sequences and Expression in *E. coli:*

The DNA coding sequence of a truncation of the *Arabidopsis thaliana* GPT coding sequence of SEQ ID NO: 1 was designed, synthesized, inserted into a PET28 vector, and expressed in *E. coli*. The DNA sequence of the truncated *Arabidopsis* GPT coding sequence used in this example is provided in SEQ ID NO: 20 (−45 AA construct), and the corresponding truncated GPT protein amino acid sequence is provided in SEQ ID NO: 21. Briefly, *E. coli* cells were

transformed with the expression vector and transformants grown overnight in LB broth diluted and grown to OD 0.4, expression induced with isopropyl-B-D-thiogalactoside (0.4 micromolar), grown for 3 hr and harvested. A total of 25×10^6 cells were then assayed for biological activity using HPLC as described in Example 6. Untransformed, wild type *E. coli* cells were assayed as a control. An additional control used *E coli* cells transformed with an empty vector.

Expression of the truncated −45 *Arabidopsis* GPT coding sequence of SEQ ID NO: 20 resulted in the over-expression of biologically active recombinant GPT protein (2-oxoglutaramate synthesis-catalyzing bioactivity). Specifically, 16.1 nanomoles of 2-oxoglutaramate activity was observed in the *E. coli* cells overexpressing the truncated −45 GPT, compared to only 0.02 nanomoles of 2-oxoglutaramate activity in control *E. coli* cells, a more than 800-fold activity level increase over control. For comparison, the full length *Arabidopsis* gene coding sequence expressed in the same *E. coli* assay generated 2.8 nanomoles of 2-oxoglutaramate activity, or roughly less than one-fifth the activity observed from the truncated recombinant GPT protein.

All publications, patents, and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

The present invention is not to be limited in scope by the embodiments disclosed herein, which are intended as single illustrations of individual aspects of the invention, and any which are functionally equivalent are within the scope of the invention. Various modifications to the models and methods of the invention, in addition to those described herein, will become apparent to those skilled in the art from the foregoing description and teachings, and are similarly intended to fall within the scope of the invention. Such modifications or other embodiments can be practiced without departing from the true scope and spirit of the invention.

TABLE OF SEQUENCES

SEQ ID NO: 1

Arabidopsis glutamine phenylpyruvate transaminase

DNA coding sequence:

ATGTACCTGGACATAAATGGTGTGATGATCAAACAGTTTAGCTTCAAAGC
CTCTCTTCTCCCATTCTCTTCTAATTTCCGACAAAGCTCCGCCAAAATCC
ATCGTCCTATCGGAGCCACCATGACCACAGTTTCGACTCAGAACGAGTCT
ACTCAAAAACCCGTCCAGGTGGCGAAGAGATTAGAGAAGTTCAAGACTAC
TATTTTCACTCAAATGAGCATATTGGCAGTTAAACATGGAGCGATCAATT
TAGGCCAAGGCTTTCCCAATTTCGACGGTCCTGATTTTGTTAAAGAAGCT
GCGATCCAAGCTATTAAAGATGGTAAAAACCAGTATGCTCGTGGATACGG

CATTCCTCAGCTCAACTCTGCTATAGCTGCGCGGTTTCGTGAAGATACGG

GTCTTGTTGTTGATCCTGAGAAAGAAGTTACTGTTACATCTGGTTGCACA

GAAGCCATAGCTGCAGCTATGTTGGGTTTAATAAACCCTGGTGATGAAGT

CATTCTCTTTGCACCGTTTTATGATTCCTATGAAGCAACACTCTCTATGG

CTGGTGCTAAAGTAAAAGGAATCACTTTACGTCCACCGGACTTCTCCATC

CCTTTGGAAGAGCTTAAAGCTGCGGTAACTAACAAGACTCGAGCCATCCT

TATGAACACTCCGCACAACCCGACCGGGAAGATGTTCACTAGGGAGGAGC

TTGAAACCATTGCATCTCTCTGCATTGAAAACGATGTGCTTGTGTTCTCG

GATGAAGTATACGATAAGCTTGCGTTTGAAATGGATCACATTTCTATAGC

TTCTCTTCCCGGTATGTATGAAAGAACTGTGACCATGAATTCCCTGGGAA

AGACTTTCTCTTTAACCGGATGGAAGATCGGCTGGGCGATTGCGCCGCCT

CATCTGACTTGGGGAGTTCGACAAGCACACTCTTACCTCACATTCGCCAC

ATCAACACCAGCACAATGGGCAGCCGTTGCAGCTCTCAAGGCACCAGAGT

CTTACTTCAAAGAGCTGAAAAGAGATTACAATGTGAAAAAGGAGACTCTG

GTTAAGGGTTTGAAGGAAGTCGGATTTACAGTGTTCCCATCGAGCGGGAC

TTACTTTGTGGTTGCTGATCACACTCCATTTGGAATGGAGAACGATGTTG

CTTTCTGTGAGTATCTTATTGAAGAAGTTGGGGTCGTTGCGATCCCAACG

AGCGTCTTTTATCTGAATCCAGAAGAAGGGAAGAATTTGGTTAGGTTTGC

GTTCTGTAAAGACGAAGAGACGTTGCGTGGTGCAATTGAGAGGATGAAGC

AGAAGCTTAAGAGAAAAGTCTGA

SEQ ID NO: 2

Arabidopsis GPT amino acid sequence

MYLDINGVMIKQFSFKASLLPFSSNFRQSSAKIHRPIGATMTTVSTQNES

TQKPVQVAKRLEKFKTTIFTQMSILAVKHGAINLGQGFPNFDGPDFVKEA

AIQAIKDGKNQYARGYGIPQLNSAIAARFREDTGLWDPEKEVTVTSGCTE

AIAAAMLGLINPGDEVILFAPFYDSYEATLSMAGAKVKGITLRPPDFSIP

LEELKAAVTNKTRAILMNTPHNPTGKMFTREELETIASLCIENDVLVFSD

EVYDKLAFEMDHISIASLPGMYERTVTMNSLGKTFSLTGWKIGWAIAPPH

LTWGVRQAHSYLTFATSTPAQWAAVAALKAPESYFKELKRDYNVKKETLV

KGLKEVGFTVFPSSGTYFWADHTPFGMENDVAFCEYLIEEVGVVAIPTSV

FYLNPEEGKNLVRFAFCKDEETLRGAIERMKQKLKRK

SEQ ID NO: 3

Grape GPT DNA sequence

Showing Cambia 1305.1 with (3′ end of) rbcS3C + Vifis (Grape). Bold **ATG** is the start site, parentheses are the catI intron and the underlined actagt is the speI cloning site used to splice in the hordeum gene.

AAAAAAGAAAAAAAAAACATATCTTGTTTGTCAGTATGGGAAGTTTGAGA

TAAGGACGAGTGAGGGGTTAAAATTCAGTGGCCATTGATTTTGTAATGCC

AAGAACCACAAAATCCAATGGTTACCATTCCTGTAAGATGAGGTTTGCTA

ACTCTTTTTGTCCGTTAGATAGGAAGCCTTATCACTATATATACAAGGCG

TCCTAATAACCTCTTAGTAACCAATTATTTCAGCA

ATGG

TAGATCTGA

GG(GTAAATTTCTAGTTTTTCTCCTTCATTTTCTTGGTTAGGACCCTTTT

CTCTTTTTATTTTTTTGAGCTTTGATCTTTCTTTAAACTGATCTATTTTT

TAATTGATTGGTTATGGTGTAAATATTACATAGCTTTAACTGATAATCTG

ATTACTTTATTTCGTGTGTCTATGATGATGATGATAGTTACAG)AACCGA

CGA

ACTAGT

ATGCAGCTCTCTCAATGTACCTGGACATTCCCAGAGTTGCT

TAAAAGACCAGCCTTTTTAAGGAGGAGTATTGATAGTATTTCGAGTAGAA

GTAGGTCCAGCTCCAAGTATCCATCTTTCATGGCGTCCGCATCAACGGTC

TCCGCTCCAAATACGGAGGCTGAGCAGACCCATAACCCCCCTCAACCTCT

ACAGGTTGCAAAGCGCTTGGAGAAATTCAAAACAACAATCTTTACTCAAA

TGAGCATGCTTGCCATCAAACATGGAGCAATAAACCTTGGCCAAGGGTTT

CCCAACTTTGATGGTCCTGAGTTTGTCAAAGAAGCAGCAATTCAAGCCAT

TAAGGATGGGAAAAACCAATATGCTCGTGGATATGGAGTTCCTGATCTCA

ACTCTGCTGTTGCTGATAGATTCAAGAAGGATACAGGACTCGTGGTGGAC

CCCGAGAAGGAAGTTACTGTTACTTCTGGATGTACAGAAGCAATTGCTGC

TACTATGCTAGGCTTGATAAATCCTGGTGATGAGGTGATCCTCTTTGCTC

CATTTTATGATTCCTATGAAGCCACTCTATCCATGGCTGGTGCCCAAATA

AAATCCATCACTTTACGTCCTCCGGATTTTGCTGTGCCCATGGATGAGCT

CAAGTCTGCAATCTCAAAGAATACCCGTGCAATCCTTATAAACACTCCCC

ATAACCCCACAGGAAAGATGTTCACAAGGGAGGAACTGAATGTGATTGCA

TCCCTCTGCATTGAGAATGATGTGTTGGTGTTTACTGATGAAGTTTACGA

CAAGTTGGCTTTCGAAATGGATCACATTTCCATGGCTTCTCTTCCTGGGA

TGTACGAGAGGACCGTGACTATGAATTCCTTAGGGAAAACTTTCTCCCTG

ACTGGATGGAAGATTGGTTGGACAGTAGCTCCCCCACACCTGACATGGGG

AGTGAGGCAAGCCCACTCATTCCTCACGTTTGCTACCTGCACCCCAATGC

AATGGGCAGCTGCAACAGCCCTCCGGGCCCCAGACTCTTACTATGAAGAG

CTAAAGAGAGATTACAGTGCAAAGAAGGCAATCCTGGTGGAGGGATTGAA

GGCTGTCGGTTTCAGGGTATACCCATCAAGTGGGACCTATTTTGTGGTGG

TGGATCACACCCCATTTGGGTTGAAAGACGATATTGCGTTTTGTGAGTAT

CTGATCAAGGAAGTTGGGGTGGTAGCAATTCCGACAAGCGTTTTCTACTT

ACACCCAGAAGATGGAAAGAACCTTGTGAGGTTTACCTTCTGTAAAGACG

AGGGAACTCTGAGAGCTGCAGTTGAAAGGATGAAGGAGAAACTGAAGCCT

AAACAATAGGGGCACGTGA

SEQ ID NO: 4

Grape GPT amino acid sequence

MVDLRNRRTSMQLSQCTWTFPELLKRPAFLRRSIDSISSRRSSSKYPSFM

ASASTVSAPNTEAEQTHNPPQPLQVAKRLEKFKTTIFTQMSMLAIKHGAI

NLGQGFPNFDGPEFVKEAAIQAIKDGKNQYARGYGVPDLNSAVADRFKKD

TGLVVDPEKEVTVTSGCTEAIAATMLGLINPGDEVILFAPFYDSYEATLS

MAGAQIKSITLRPPDFAVPMDELKSAISKNTRAILINTPHNPTGKMFTRE

ELNVIASLCIENDVLVFTDEVYDKLAFEMDHISMASLPGMYERTVTMNSL

GKTFSLTGWKIGWTVAPPHLTWGVRQAHSFLTFATCTPMQWAAATALRAP

DSYYEELKRDYSAKKAILVEGLKAVGFRVYPSSGTYFVVVDHTPFGLKDD

IAFCEYLIKEVGVVAIPTSVFYLHPEDGKNLVRFTFCKDEGTLRAAVERM

KEKLKPKQ

SEQ ID NO: 5

Rice GPT DNA sequence

Rice GPT codon optimized for *E. coli* expression;

untranslated sequences shown in lower case

atgtggATGAACCTGGCAGGCTTTCTGGCAACCCCGGCAACCGCAACCGC

AACCCGTCATGAAATGCCGCTGAACCCGAGCAGCAGCGCGAGCTTTCTGC

TGAGCAGCCTGCGTCGTAGCCTGGTGGCGAGCCTGCGTAAAGCGAGCCCG

GCAGCAGCAGCAGCACTGAGCCCGATGGCAAGCGCAAGCACCGTGGCAGC

AGAAAACGGTGCAGCAAAAGCAGCAGCAGAAAAACAGCAGCAGCAGCCGG

TGCAGGTGGCGAAACGTCTGGAAAAATTTAAAACCACCATTTTTACCCAG

ATGAGCATGCTGGCGATTAAACATGGCGCGATTAACCTGGGCCAGGGCTT

TCCGAACTTTGATGGCCCGGATTTTGTGAAAGAAGCGGCGATTCAGGCGA

TTAACGCGGGCAAAAACCAGTATGCGCGTGGCTATGGCGTGCCGGAACTG

AACAGCGCGATTGCGGAACGTTTTCTGAAAGATAGCGGCCTGCAGGTGGA

TCCGGAAAAAGAAGTGACCGTGACCAGCGGCTGCACCGAAGCGATTGCGG

CGACCATTCTGGGCCTGATTAACCCGGGCGATGAAGTGATTCTGTTTGCG

CCGTTTTATGATAGCTATGAAGCGACCCTGAGCATGGCGGGCGCGAACGT

GAAAGCGATTACCCTGCGTCCGCCGGATTTTAGCGTGCCGCTGGAAGAAC

TGAAAGCGGCCGTGAGCAAAAACACCCGTGCGATTATGATTAACACCCCG

CATAACCCGACCGGCAAAATGTTTACCCGTGAAGAACTGGAATTTATTGC

GACCCTGTGCAAAGAAAACGATGTGCTGCTGTTTGCGGATGAAGTGTATG

ATAAACTGGCGTTTGAAGCGGATCATATTAGCATGGCGAGCATTCCGGGC

ATGTATGAACGTACCGTGACCATGAACAGCCTGGGCAAAACCTTTAGCCT

GACCGGCTGGAAAATTGGCTGGGCGATTGCGCCGCCGCATCTGACCTGGG

GCGTGCGTCAGGCACATAGCTTTCTGACCTTTGCAACCTGCACCCCGATG

CAGGCAGCCGCCGCAGCAGCACTGCGTGCACCGGATAGCTATTATGAAGA

ACTGCGTCGTGATTATGGCGCGAAAAAAGCGCTGCTGGTGAACGGCCTGA

AAGATGCGGGCTTTATTGTGTATCCGAGCAGCGGCACCTATTTTGTGATG

GTGGATCATACCCCGTTTGGCTTTGATAACGATATTGAATTTTGCGAATA

TCTGATTCGTGAAGTGGGCGTGGTGGCGATTCCGCCGAGCGTGTTTTATC

TGAACCCGGAAGATGGCAAAAACCTGGTGCGTTTTACCTTTTGCAAAGAT

GATGAAACCCTGCGTGCGGCGGTGGAACGTATGAAAACCAAACTGCGTAA

AAAAAAGCTTgcggccgcactcgagcaccaccaccaccaccactga

SEQ ID NO: 6

Rice GPT amino acid sequence

Includes amino terminal amino acids MW for cloning

and His tag sequences from pet28 vector in

italics.

*MW*MNLAGFLATPATATATRHEMPLNPSSSASFLLSSLRRSLVASLRKASP

AAAAALSPMASASTVAAENGAAKAAAEKQQQQPVQVAKRLEKFKTTIFTQ

MSMLAIKHGAINLGQGFPNFDGPDFVKEAAIQAINAGKNQYARGYGVPEL

NSAIAERFLKDSGLQVDPEKEVTVTSGCTEAIAATILGLINPGDEVILFA

PFYDSYEATLSMAGANVKAITLRPPDFSVPLEELKAAVSKNTRAIMINTP

HNPTGKMFTREELEFIATLCKENDVLLFADEVYDKLAFEADHISMASIPG

MYERTVTMNSLGKTFSLTGWKIGWAIAPPHLTWGVRQAHSFLTFATCTPM

QAAAAAALRAPDSYYEELRRDYGAKKALLVNGLKDAGFIVYPSSGTYFVM

VDHTPFGFDNDIEFCEYLIREVGVVAIPPSVFYLNPEDGKNLVRFTFCKD

DETLRAAVERMKTKLRK*KKLAAALEHHHHHH*

SEQ ID NO: 7

Soybean GPT DNA sequence

TOPO 151D WITH SOYBEAN for *E coli* expression

From starting codon. Vector sequences are italicized

ATG *CATCATCACCATCACCATGGTAAGCCTATCCCTAACCCTCTCCTCGG*
*TCTCGATTCTACGGAAAACCTGTATTTTCAGGGAATTGATCCCTTCACC*G
CGAAACGTCTGGAAAAATTTCAGACCACCATTTTTACCCAGATGAGCCTG
CTGGCGATTAAACATGGCGCGATTAACCTGGGCCAGGGCTTTCCGAACTT
TGATGGCCCGGAATTTGTGAAAGAAGCGGCGATTCAGGCGATTCGTGATG
GCAAAAACCAGTATGCGCGTGGCTATGGCGTGCCGGATCTGAACATTGCG
ATTGCGGAACGTTTTAAAAAAGATACCGGCCTGGTGGTGGATCCGGAAAA
AGAAATTACCGTGACCAGCGGCTGCACCGAAGCGATTGCGGCGACCATGA
TTGGCCTGATTAACCCGGGCGATGAAGTGATTATGTTTGCGCCGTTTTAT
GATAGCTATGAAGCGACCCTGAGCATGGCGGGCGCGAAAGTGAAAGGCAT
TACCCTGCGTCCGCCGGATTTTGCGGTGCCGCTGGAAGAACTGAAAAGCA
CCATTAGCAAAAACACCCGTGCGATTCTGATTAACACCCCGCATAACCCG
ACCGGCAAAATGTTTACCCGTGAAGAACTGAACTGCATTGCGAGCCTGTG
CATTGAAAACGATGTGCTGGTGTTTACCGATGAAGTGTATGATAAACTGG
CGTTTGATATGGAACATATTAGCATGGCGAGCCTGCCGGGCATGTTTGAA
CGTACCGTGACCCTGAACAGCCTGGGCAAAACCTTTAGCCTGACCGGCTG
GAAAATTGGCTGGGCGATTGCGCCGCCGCATCTGAGCTGGGGCGTGCGTC
AGGCGCATGCGTTTCTGACCTTTGCAACCGCACATCCGTTTCAGTGCGCA
GCAGCAGCAGCACTGCGTGCACCGGATAGCTATTATGTGGAACTGAAACG
TGATTATATGGCGAAACGTGCGATTCTGATTGAAGGCCTGAAAGCGGTGG
GCTTTAAAGTGTTTCCGAGCAGCGGCACCTATTTTGTGGTGGTGGATCAT
ACCCCGTTTGGCCTGGAAAACGATGTGGCGTTTTGCGAATATCTGGTGAA
AGAAGTGGGCGTGGTGGCGATTCCGACCAGCGTGTTTTATCTGAACCCGG
AAGAAGGCAAAAACCTGGTGCGTTTTACCTTTTGCAAAGATGAAGAAACC
ATTCGTAGCGCGGTGGAACGTATGAAAGCGAAACTGCGTAAAGTCGACTA
A

SEQ ID NO: 8

Soybean GPT amino acid sequence

Translated protein product, vector sequences

italicized

*MHHHHHHGKPIPNPLLGLDSTENLYFQGIDPFT*AKRLEKFQTTIFTQMSL

LAIKHGAINLGQGFPNFDGPEFVKEAAIQAIRDGKNQYARGYGVPDLNIA

IAERFKKDTGLWDPEKEITVTSGCTEAIAATMIGLINPGDEVIMFAPFYD

SYEATLSMAGAKVKGITLRPPDFAVPLEELKSTISKNTRAILINTPHNPT

GKMFTREELNCIASLCIENDVLVFTDEVYDKLAFDMEHISMASLPGMFER

TVTINSLGKTFSLTGWKIGWAIAPPHLSWGVRQAHAFLTFATAHPFQCAA

AAALRAPDSYYVELKRDYMAKRAILIEGLKAVGFKVFPSSGTYFVVVDHT

PFGLENDVAFCEYLVKEVGVVAIPTSVFYLNPEEGKNLVRFTFCKDEETI

RSAVERMKAKLRKVD

SEQ ID NO: 9

Barley GPT DNA sequence

Coding sequence from start with intron removed

ATGG

TAGATCTGAGGAACCGACGA

ACTAGT

ATGGCATCCGCCCCCGCCTC

CGCCTCCGCGGCCCTCTCCACCGCCGCCCCCGCCGACAACGGGGCCGCCA

AGCCCACGGAGCAGCGGCCGGTACAGGTGGCTAAGCGATTGGAGAAGTTC

AAAACAACAATTTTCACACAGATGAGCATGCTCGCAGTGAAGCATGGAGC

AATAAACCTTGGACAGGGGTTTCCCAATTTTGATGGCCCTGACTTTGTCA

AAGATGCTGCTATTGAGGCTATCAAAGCTGGAAAGAATCAGTATGCAAGA

GGATATGGTGTGCCTGAATTGAACTCAGCTGTTGCTGAGAGATTTCTCAA

GGACAGTGGATTGCACATCGATCCTGATAAGGAAGTTACTGTTACATCTG

GGTGCACAGAAGCAATAGCTGCAACGATATTGGGTCTGATCAACCCTGGG

GATGAAGTCATACTGTTTGCTCCATTCTATGATTCTTATGAGGCTACACT

GTCCATGGCTGGTGCGAATGTCAAAGCCATTACACTCCGCCCTCCGGACT

TTGCAGTCCCTCTTGAAGAGCTAAAGGCTGCAGTCTCGAAGAATACCAGA

GCAATAATGATTAATACACCTCACAACCCTACCGGGAAAATGTTCACAAG

GGAGGAACTTGAGTTCATTGCTGATCTCTGCAAGGAAAATGACGTGTTGC

TCTTTGCCGATGAGGTCTACGACAAGCTGGCGTTTGAGGCGGATCACATA

TCAATGGCTTCTATTCCTGGCATGTATGAGAGGACCGTCACTATGAACTC

CCTGGGGAAGACGTTCTCCTTGACCGGATGGAAGATCGGCTGGGCGATAG

CACCACCGCACCTGACATGGGGCGTAAGGCAGGCACACTCCTTCCTCACA

TTCGCCACCTCCACGCCGATGCAATCAGCAGCGGCGGCGGCCCTGAGAGC

ACCGGACAGCTACTTTGAGGAGCTGAAGAGGGACTACGGCGCAAAGAAAG

CGCTGCTGGTGGACGGGCTCAAGGCGGCGGGCTTCATCGTCTACCCTTCG

AGCGGAACCTACTTCATCATGGTCGACCACACCCCGTTCGGGTTCGACAA

CGACGTCGAGTTCTGCGAGTACTTGATCCGCGAGGTCGGCGTCGTGGCCA

TCCCGCCAAGCGTGTTCTACCTGAACCCGGAGGACGGGAAGAACCTGGTG

AGGTTCACCTTCTGCAAGGACGACGACACGCTAAGGGCGGCGGTGGACAG

GATGAAGGCCAAGCTCAGGAAGAAATGA

SEQ ID NO: 10

Barley GPT amino acid sequence

Translated sequence from start site (intron removed)

MVDLRNRRTSMASAPASASAALSTAAPADNGAAKPTEQRPVQVAKRLEKF

KTTIFTQMSMLAVKHGAINLGQGFPNFDGPDFVKDAAIEAIKAGKNQYAR

GYGVPELNSAVAERFLKDSGLHIDPDKEVTVTSGCTEAIAATILGLINPG

DEVILFAPFYDSYEATLSMAGANVKAITLRPPDFAVPLEELKAAVSKNTR

AIMINTPHNPTGKMFTREELEFIADLCKENDVLLFADEVYDKLAFEADHI

SMASIPGMYERTVTMNSLGKTFSLTGWKIGWAIAPPHLTWGVRQAHSFLT

FATSTPMQSAAAAALRAPDSYFEELKRDYGAKKALLVDGLKAAGFIVYPS

SGTYFIMVDHTPFGFDNDVEFCEYLIREVGWAIPPSVFYLNPEDGKNLVR

FTFCKDDDTLRAAVDRMKAKLRKK

SEQ ID NO: 11

Zebra fish GPT DNA sequence

Danio rerio sequence designed for expression in *E coli.* Bold, italicized nucleotides added for cloning or from pET28b vector.

ATGTCC

GTGGCGAAACGTCTGGAAAAATTTAAAACCACCATTTTTACCCA

GATGAGCATGCTGGCGATTAAACATGGCGCGATTAACCTGGGCCAGGGCT

TTCCGAACTTTGATGGCCCGGATTTTGTGAAAGAAGCGGCGATTCAGGCG

ATTCGTGATGGCAACAACCAGTATGCGCGTGGCTATGGCGTGCCGGATCT

GAACATTGCGATTAGCGAACGTTATAAAAAAGATACCGGCCTGGCGGTGG

ATCCGGAAAAAGAAATTACCGTGACCAGCGGCTGCACCGAAGCGATTGCG

GCGACCGTGCTGGGCCTGATTAACCCGGGCGATGAAGTGATTGTGTTTGC

GCCGTTTTATGATAGCTATGAAGCGACCCTGAGCATGGCGGGCGCGAAAG

TGAAAGGCATTACCCTGCGTCCGCCGGATTTTGCGCTGCCGATTGAAGAA

CTGAAAAGCACCATTAGCAAAAACACCCGTGCGATTCTGCTGAACACCCC

GCATAACCCGACCGGCAAAATGTTTACCCCGGAAGAACTGAACACCATTG

CGAGCCTGTGCATTGAAAACGATGTGCTGGTGTTTAGCGATGAAGTGTAT

GATAAACTGGCGTTTGATATGGAACATATTAGCATTGCGAGCCTGCCGGG

CATGTTTGAACGTACCGTGACCATGAACAGCCTGGGCAAAACCTTTAGCC

TGACCGGCTGGAAAATTGGCTGGGCGATTGCGCCGCCGCATCTGACCTGG

GGCGTGCGTCAGGCGCATGCGTTTCTGACCTTTGCAACCAGCAACCCGAT

GCAGTGGGCAGCAGCAGTGGCACTGCGTGCACCGGATAGCTATTATACCG

AACTGAAACGTGATTATATGGCGAAACGTAGCATTCTGGTGGAAGGCCTG

AAAGCGGTGGGCTTTAAAGTGTTTCCGAGCAGCGGCACCTATTTTGTGGT

GGTGGATCATACCCCGTTTGGCCATGAAAACGATATTGCGTTTTGCGAAT

ATCTGGTGAAAGAAGTGGGCGTGGTGGCGATTCCGACCAGCGTGTTTTAT

CTGAACCCGGAAGAAGGCAAAAACCTGGTGCGTTTTACCTTTTGCAAAGA

TGAAGGCACCCTGCGTGCGGCGGTGGATCGTATGAAAGAAAAACTGCGT

GTCGACAAGC

CTTGCGGC

CGCACTCG

AGCACCAC

CACCACCA

CCACTGA

SEQ ID NO: 12

Zebra fish GPR amino acid sequence

Amino acid sequence of *Danio rerio* cloned and expressed in *E. coli* (bold, italicized amino acids are added from vector/ cloning and His tag on C-terminus)

MS VAKRLEKFKTTIFTQMSMLAIKHGAINLGQGFPNFDGPDFVKEAAIQA

IRDGNNQYARGYGVPDLNIAISERYKKDTGLAVDPEKEITVTSGCTEAIA

ATVLGLINPGDEVIVFAPFYDSYEATLSMAGAKVKGITLRPPDFALPIEE

LKSTISKNTRAILLNTPHNPTGKMFTPEELNTIASLCIENDVLVFSDEVY

DKLAFDMEHISIASLPGMFERTVTMNSLGKTFSLTGWKIGWAIAPPHLTW

GVRQAHAFLTFATSNPMQWAAAVALRAPDSYYTELKRDYMAKRSILVEGL

KAVGFKVFPSSGTYFVVVDHTPFGHENDIAFCEYLVKEVGVVAIPTSVFY

LNPEEGKNLVRFTFCKDEGTLRAAVDRMKEKLRK

VDKLAAAL

EHHHHHH-

SEQ ID NO: 13

Arabidopsis truncated GPT −30 construct DNA sequence *Arabidopsis* GPT with 30 amino acids removed from the targeting sequence.

ATGGCCAAAATCCATCGTCCTATCGGAGCCACCATGACCACAGTTTCGAC

TCAGAACGAGTCTACTCAAAAACCCGTCCAGGTGGCGAAGAGATTAGAGA

AGTTCAAGACTACTATTTTCACTCAAATGAGCATATTGGCAGTTAAACAT

GGAGCGATCAATTTAGGCCAAGGCTTTCCCAATTTCGACGGTCCTGATTT

TGTTAAAGAAGCTGCGATCCAAGCTATTAAAGATGGTAAAAACCAGTATG

CTCGTGGATACGGCATTCCTCAGCTCAACTCTGCTATAGCTGCGCGGTTT

CGTGAAGATACGGGTCTTGTTGTTGATCCTGAGAAAGAAGTTACTGTTAC

ATCTGGTTGCACAGAAGCCATAGCTGCAGCTATGTTGGGTTTAATAAACC

CTGGTGATGAAGTCATTCTUTTGCACCGTTTTATGATTCCTATGAAGCAA

CACTCTCTATGGCTGGTGCTAAAGTAAAAGGAATCACTTTACGTCCACCG

GACTTCTCCATCCCTTTGGAAGAGCTTAAAGCTGCGGTAACTAACAAGAC

TCGAGCCATCCTTATGAACACTCCGCACAACCCGACCGGGAAGATGTTCA

CTAGGGAGGAGCTTGAAACCATTGCATCTCTCTGCATTGAAAACGATGTG

CTTGTGTTCTCGGATGAAGTATACGATAAGCTTGCGTTTGAAATGGATCA

CATTTCTATAGCTTCTCTTCCCGGTATGTATGAAAGAACTGTGACCATGA

ATTCCCTGGGAAAGACTTTCTCTTTAACCGGATGGAAGATCGGCTGGGCG

ATTGCGCCGCCTCATCTGACTTGGGGAGTTCGACAAGCACACTCTTACCT

CACATTCGCCACATCAACACCAGCACAATGGGCAGCCGTTGCAGCTCTCA

AGGCACCAGAGTCTTACTTCAAAGAGCTGAAAAGAGATTACAATGTGAAA

AAGGAGACTCTGGTTAAGGGTTTGAAGGAAGTCGGATTTACAGTGTTCCC

ATCGAGCGGGACTTACTTTGTGGTTGCTGATCACACTCCATTTGGAATGG

AGAACGATGTTGCTTTCTGTGAGTATCTTATTGAAGAAGTTGGGGTCGTT

GCGATCCCAACGAGCGTCTTTTATCTGAATCCAGAAGAAGGGAAGAATTT

GGTTAGGTTTGCGTTCTGTAAAGACGAAGAGACGTTGCGTGGTGCAATTG

AGAGGATGAAGCAGAAGCTTAAGAGAAAAGTCTGA

SEQ ID NO: 14

Arabidopsis truncated GPT −30 construct amino acid sequence

MAKIHRPIGATMTTVSTQNESTQKPVQVAKRLEKFKTTIFTQMSILAVKH

GAINLGQGFPNFDGPDFVKEAAIQAIKDGKNQYARGYGIPQLNSAIAARF

REDTGLVVDPEKEVTVTSGCTEAIAAAMLGLINPGDEVILFAPFYDSYEA

TLSMAGAKVKGITLRPPDFSIPLEELKAAVTNKTRAILMNTPHNPTGKMF

TREELETIASLCIENDVLVFSDEVYDKLAFEMDHISIASLPGMYERTVTM

NSLGKTFSLTGWKIGWAIAPPHLTWGVRQAHSYLTFATSTPAQWAAVAAL

KAPESYFKELKRDYNVKKETLVKGLKEVGFTVFPSSGTYFVVADHTPFGM

ENDVAFCEYLIEEVGVVAIPTSVFYLNPEEGKNLVRFAFCKDEETLRGAI

ERMKQKLKRKV

SEQ ID NO: 15:

Arabidopsis truncated GPT −45 construct DNA sequence Arabidopsis GPT with 45 residues in the targeting sequence removed

ATGGCGACTCAGAACGAGTCTACTCAAAAACCCGTCCAGGTGGCGAAGAG

ATTAGAGAAGTTCAAGACTACTATTTTCACTCAAATGAGCATATTGGCAG

TTAAACATGGAGCGATCAATTTAGGCCAAGGCTTTCCCAATTTCGACGGT

CCTGATTTTGTTAAAGAAGCTGCGATCCAAGCTATTAAAGATGGTAAAAA

CCAGTATGCTCGTGGATACGGCATTCCTCAGCTCAACTCTGCTATAGCTG

CGCGGTTTCGTGAAGATACGGGTCTTGTTGTTGATCCTGAGAAAGAAGTT

ACTGTTACATCTGGTTGCACAGAAGCCATAGCTGCAGCTATGTTGGGTTT

AATAAACCCTGGTGATGAAGTCATTCTCTTTGCACCGTTTTATGATTCCT

ATGAAGCAACACTCTCTATGGCTGGTGCTAAAGTAAAAGGAATCACTTTA

CGTCCACCGGACTTCTCCATCCCTTTGGAAGAGCTTAAAGCTGCGGTAAC

TAACAAGACTCGAGCCATCCTTATGAACACTCCGCACAACCCGACCGGGA

AGATGTTCACTAGGGAGGAGCTTGAAACCATTGCATCTCTCTGCATTGAA

AACGATGTGCTTGTGTTCTCGGATGAAGTATACGATAAGCTTGCGTTTGA

AATGGATCACATTTCTATAGCTTCTCTTCCCGGTATGTATGAAAGAACTG

TGACCATGAATTCCCTGGGAAAGACTTTCTCTTTAACCGGATGGAAGATC

GGCTGGGCGATTGCGCCGCCTCATCTGACTTGGGGAGTTCGACAAGCACA

CTCTTACCTCACATTCGCCACATCAACACCAGCACAATGGGCAGCCGTTG

CAGCTCTCAAGGCACCAGAGTCTTACTTCAAAGAGCTGAAAAGAGATTAC

AATGTGAAAAAGGAGACTCTGGTTAAGGGTTTGAAGGAAGTCGGATTTAC

AGTGTTCCCATCGAGCGGGACTTACTTTGTGGTTGCTGATCACACTCCAT
TTGGAATGGAGAACGATGTTGCTTTCTGTGAGTATCTTATTGAAGAAGTT

GGGGTCGTTGCGATCCCAACGAGCGTCTTTTATCTGAATCCAGAAGAAGG
GAAGAATTTGGTTAGGTTTGCGTTCTGTAAAGACGAAGAGACGTTGCGTG
GTGCAATTGAGAGGATGAAGCAGAAGCTTAAGAGAAAAGTCTGA

SEQ ID NO: 16:

Arabidopsis truncated GPT −45 construct amino acid sequence

MATQNESTQKPVQVAKRLEKFKTTIFTQMSILAVKHGAINLGQGFPNFDG
PDFVKEAAIQAIKDGKNQYARGYGIPQLNSAIAARFREDTGLVVDPEKEV

TVTSGCTEAIAAAMLGLINPGDEVILFAPFYDSYEATLSMAGAKVKGITL
RPPDFSIPLEELKAAVTNKTRAILMNTPHNPTGKMFTREELETIASLCIE
NDVLVFSDEVYDKLAFEMDHISIASLPGMYERTVTMNSLGKTFSLTGWKI
GWAIAPPHLTWGVRQAHSYLTFATSTPAQWAAVAALKAPESYFKELKRDY
NVKKETLVKGLKEVGFTVFPSSGTYFVVADHTPFGMENDVAFCEYLIEEV
GVVAIPTSVFYLNPEEGKNLVRFAFCKDEETLRGAIERMKQKLKRKV

SEQ ID NO: 17:

Tomato Rubisco promoter TOMATO RuBisCo rbcS3C promoter sequence from KpnI to NcoI

*GGTACC*GTTTGAATCCTCCITAAAGTTTTTCTCTGGAGAAACTGTAGTAA
TTTTACTTTGTTGTGTTCCCTTCATCTTTTGAATTAATGGCATTTGTTTT
AATACTAATCTGCTTCTGAAACTTGTAATGTATGTATATCAGTTTCTTAT
AATTTATCCAAGTAATATCTTCCATTCTCTATGCAATTGCCTGCATAAGC
TCGACAAAAGAGTACATCAACCCCTCCTCCTCTGGACTACTCTAGCTAAA
CTTGAATTTCCCCTTAAGATTATGAAATTGATATATCCTTAACAAACGAC
TCCTTCTGTTGGAAAATGTAGTACTTGTCTTTCTTCTTTTGGGTATATAT
AGTTTATATACACCATACTATGTACAACATCCAAGTAGAGTGAAATGGAT
ACATGTACAAGACTTATTTGATTGATTGATGACTTGAGTTGCCTTAGGAG

TAACAAATTCTTAGGTCAATAAATCGTTGATTTGAAATTAATCTCTCTGT

CTTAGACAGATAGGAATTATGACTTCCAATGGTCCAGAAAGCAAAGTTCG

CACTGAGGGTATACTTGGAATTGAGACTTGCACAGGTCCAGAAACCAAAG

TTCCCATCGAGCTCTAAAATCACATCTTTGGAATGAAATTCAATTAGAGA

TAAGTTGCTTCATAGCATAGGTAAAATGGAAGATGTGAAGTAACCTGCAA

TAATCAGTGAAATGACATTAATACACTAAATACTTCATATGTAATTATCC

TTTCCAGGTTAACAATACTCTATAAAGTAAGAATTATCAGAAATGGGCTC

ATCAAACTTTTGTACTATGTATTTCATATAAGGAAGTATAACTATACATA

AGTGTATACACAACTTTATTCCTATTTTGTAAAGGTGGAGAGACTGTTTT

CGATGGATCTAAAGCAATATGTCTATAAAATGCATTGATATAATAATTAT

CTGAGAAAATCCAGAATTGGCGTTGGATTATTTCAGCCAAATAGAAGTTT

GTACCATACTTGTTGATTCCTTCTAAGTTAAGGTGAAGTATCATTCATAA

ACAGTTTTCCCCAAAGTACTACTCACCAAGTTTCCCTTTGTAGAATTAAC

AGTTCAAATATATGGCGCAGAAATTACTCTATGCCCAAAACCAAACGAGA

AAGAAACAAAATACAGGGGTTGCAGACTTTATTTTCGTGTTAGGGTGTGT

TTTTTCATGTAATTAATCAAAAAATATTATGACAAAAACATTTATACATA

TTTTTACTCAACACTCTGGGTATCAGGGTGGGTTGTGTTCGACAATCAAT

ATGGAAAGGAAGTATTTTCCTTATTTTTTTAGTTAATATTTTCAGTTATA

CCAAACATACCTTGTGATATTATTTTTAAAAATGAAAAACTCGTCAGAAA

GAAAAAGCAAAAGCAACAAAAAAATTGCAAGTATTTTTTAAAAAAGAAAA

AAAAAACATATCTTGTTTGTCAGTATGGGAAGTTTGAGATAAGGACGAGT

GAGGGGTTAAAATTCAGTGGCCATTGATTTTGTAATGCCAAGAACCACAA

AATCCAATGGTTACCATTCCTGTAAGATGAGGTTTGCTAACTCTTTTTGT

CCGTTAGATAGGAAGCCTTATCACTATATATACAAGGCGTCCTAATAACC

TCTTAGTAACCAATTATTTCAGCACC *ATG* G

SEQ ID NO: 18:

Bamboo GPT DNA sequence

ATGGCCTCCGCGGCCGTCTCCACCGTCGCCACCGCCGCCGACGGCGTCGC

GAAGCCGACGGAGAAGCAGCCGGTACAGGTCGCAAAGCGTTTGGAAAAGT

TTAAGACAACAATTTTCACACAGATGAGCATGCTTGCCATCAAGCATGGA

GCAATAAACCTCGGCCAGGGCTTTCCGAATTTTGATGGCCCTGACTTTGT

GAAAGAAGCTGCTATTCAAGCTATCAATGCTGGGAAGAATCAGTATGCAA

GAGGATATGGTGTGCCTGAACTGAACTCGGCTGTTGCTGAAAGGTTCCTG

AAGGACAGTGGCTTGCAAGTCGATCCCGAGAAGGAAGTTACTGTCACATC

TGGGTGCACGGAAGCGATAGCTGCAACGATATTGGGTCTTATCAACCCTG

GCGATGAAGTGATCTTGTTTGCTCCATTCTATGATTCATACGAGGCTACG

CTGTCGATGGCTGGTGCCAATGTAAAAGCCATTACTCTCCGTCCTCCAGA

TTTTGCAGTCCCTCTTGAGGAGCTAAAGGCCACAGTCTCTAAGAACACCA

GAGCGATAATGATAAACACACCACACAATCCTACTGGGAAAATGTTTTCT

AGGGAAGAACTTGAATTCATTGCTACTCTCTGCAAGAAAAATGATGTGTT

GCTTTTTGCTGATGAGGTCTATGACAAGTTGGCATTTGAGGCAGATCATA

TATCAATGGCTTCTATTCCTGGCATGTATGAGAGGACTGTGACTATGAAC

TCTCTGGGGAAGACATTCTCTCTAACAGGATGGAAGATCGGTTGGGCAAT

AGCACCACCACACCTGACATGGGGTGTAAGGCAGGCACACTCATTCCTCA

CATTTGCCACCTGCACACCAATGCAATCGGCGGCGGCGGCGGCTCTTAGA

GCACCAGATAGCTACTATGGGGAGCTGAAGAGGGATTACGGTGCAAAGAA

AGCGATACTAGTCGACGGACTCAAGGCTGCAGGTTTTATTGTTTACCCTT

CAAGTGGAACATACTTTGTCATGGTCGATCACACCCCGTTTGGTTTCGAC

AATGATATTGAGTTCTGCGAGTATTTGATCCGCGAAGTCGGTGTTGTCGC

CATACCACCAAGCGTATTTTATCTCAACCCTGAGGATGGGAAGAACTTGG

TGAGGTTCACCTTCTGCAAGGATGATGATACGCTGAGAGCCGCAGTTGAG

AGGATGAAGACAAAGCTCAGGAAAAAATGA

SEQ ID NO: 19:

Bamboo GPT amino acid sequence

MASAAVSTVATAADGVAKPTEKQPVQVAKRLEKFKTTIFTQMSMLAIKHG

AINLGQGFPNFDGPDFVKEAAIQAINAGKNQYARGYGVPELNSAVAERFL

KDSGLQVDPEKEVTVTSGCTEAIAATILGLINPGDEVILFAPFYDSYEAT

LSMAGANVKAITLRPPDFAVPLEELKATVSKNTRAIMINTPHNPTGKMFS

REELEFIATLCKKNDVLLFADEVYDKLAFEADHISMASIPGMYERTVTMN

SLGKTFSLTGWKIGWAIAPPHLTWGVRQAHSFLTFATCTPMQSAAAAALR

APDSYYGELKRDYGAKKAILVDGLKAAGFIVYPSSGTYFVMVDHTPFGFD

NDIEFCEYLIREVGVVAIPPSVFYLNPEDGKNLVRFTFCKDDDTLRAAVE

RMKTKLRKK

SEQ ID NO: 20:

1305.1 + rbcS3C promoter + catI intron with rice

GPT gene.

Cambia1305.1 with (3′ end of) rbcS3C + rice GPT.

Underlined ATG is start site, parentheses are the

catI intron and the underlined actagt is the spel

cloning site used to splice in the rice gene.

AAAAAAGAAAAAAAAAACATATCTTGTTTGTCAGTATGGGAAGTTTGAGA

TAAGGACGAGTGAGGGGTTAAAATTCAGTGGCCATTGATTTTGTAATGCC

AAGAACCACAAAATCCAATGGTTACCATTCCTGTAAGATGAGGTTTGCTA

ACTCTTTTTGTCCGTTAGATAGGAAGCCTTATCACTATATATACAAGGCG

TCCTAATAACCTCTTAGTAACCAATTATTTCAGCA

CCATGG

TAGATCTGA

GG(GTAAATTTCTAGTTTTTCTCCTTCATTTTCTTGGTTAGGACCCTTTT

CTCTTTTTATTTTTTTGAGCTTTGATCTTTCTTTAAACTGATCTATTTTT

TAATTGATTGGTTATGGTGTAAATATTACATAGCTTTAACTGATAATCTG

ATTACTTTATTTCGTGTGTCTATGATGATGATGATAGTTACAG)AACCGA

CGA

ACTAGT

ATGAATCTGGCCGGCTTTCTCGCCACGCCCGCGACCGCGAC

CGCGACGCGGCATGAGATGCCGTTAAATCCCTCCTCCTCCGCCTCCTTCC

TCCTCTCCTCGCTCCGCCGCTCGCTCGTCGCGTCGCTCCGGAAGGCCTCG

CCGGCGGCGGCCGCGGCGCTCTCCCCCATGGCCTCCGCGTCCACCGTCGC

CGCCGAGAACGGCGCCGCCAAGGCGGCGGCGGAGAAGCAGCAGCAGCAGC

CTGTGCAGGTTGCAAAGCGGTTGGAAAAGTTTAAGACGACCATTTTCACA

CAGATGAGTATGCTTGCCATCAAGCATGGAGCAATAAACCTTGGCCAGGG

TTTTCCGAATTTCGATGGCCCTGACTTTGTAAAAGAGGCTGCTATTCAAG

CTATCAATGCTGGGAAGAATCAGTACGCAAGAGGATATGGTGTGCCTGAA

CTGAACTCAGCTATTGCTGAAAGATTCCTGAAGGACAGCGGACTGCAAGT

CGATCCGGAGAAGGAAGTTACTGTCACATCTGGATGCACAGAAGCTATAG

CTGCAACAATTTTAGGTCTAATTAATCCAGGCGATGAAGTGATATTGTTT

GCTCCATTCTATGATTCATATGAGGCTACCCTGTCAATGGCTGGTGCCAA

CGTAAAAGCCATTACTCTCCGTCCTCCAGATTTTTCAGTCCCTCTTGAAG

AGCTAAAGGCTGCAGTCTCGAAGAACACCAGAGCTATTATGATAAACACC

CCGCACAATCCTACTGGGAAAATGTTTACAAGGGAAGAACTTGAGTTTAT

TGCCACTCTCTGCAAGGAAAATGATGTGCTGCTTTTTGCTGATGAGGTCT

ACGACAAGTTAGCTTTTGAGGCAGATCATATATCAATGGCTTCTATTCCT

GGCATGTATGAGAGGACCGTGACCATGAACTCTCTTGGGAAGACATTCTC

TCTTACAGGATGGAAGATCGGTTGGGCAATCGCACCGCCACACCTGACAT

GGGGTGTAAGGCAGGCACACTCATTCCTCACGTTTGCGACCTGCACACCA

ATGCAAGCAGCTGCAGCTGCAGCTCTGAGAGCACCAGATAGCTACTATGA

GGAACTGAGGAGGGATTATGGAGCTAAGAAGGCATTGCTAGTCAACGGAC

TCAAGGATGCAGGTTTCATTGTCTATCCTTCAAGTGGAACATACTTCGTC

ATGGTCGACCACACCCCATTTGGTTTCGACAATGATATTGAGTTCTGCGA

GTATTTGATTCGCGAAGTCGGTGTTGTCGCCATACCACCTAGTGTATTTT

ATCTCAACCCTGAGGATGGGAAGAACTTGGTGAGGTTCACCTTTTGCAAG

GATGATGAGACGCTGAGAGCCGCGGTTGAGAGGATGAAGACAAAGCTCAG

GAAAAAATGA

SEQ ID NO: 21:

HORDEUM GPT SEQUENCE IN VECTOR

Cambia1305.1 with (3′ end of) rbcS3C + hordeum

ID14. Underlined ATG is start site, parentheses are the catI intron and the underlined actagt is the speI cloning site used to splice in the hordeum gene.

AAAAAAGAAAAAAAAAACATATCTTGTTTGTCAGTATGGGAAGTTTGAGA

TAAGGACGAGTGAGGGGTTAAAATTCAGTGGCCATTGATTTTGTAATGCC

AAGAACCACAAAATCCAATGGTTACCATTCCTGTAAGATGAGGTTTGCTA

ACTCTTTTTGTCCGTTAGATAGGAAGCCTTATCACTATATATACAAGGCG

TCCTAATAACCTCTTAGTAACCAATTATTTCAGCA

CCATGG

TAGATCTGA

GG(GTAAATTTCTAGTTTTTCTCCTTCATTTTCTTGGTTAGGACCCTTTT

CTCTTTTTATTTTTTTGAGCTTTGATCTTTCTTTAAACTGATCTATTTTT

TAATTGATTGGTTATGGTGTAAATATTACATAGCTTTAACTGATAATCTG

ATTACTTTATTTCGTGTGTCTATGATGATGATGATAGTTACAG)AACCGA

CGA

ACTAGT

ATGGCATCCGCCCCCGCCTCCGCCTCCGCGGCCCTCTCCAC

CGCCGCCCCCGCCGACAACGGGGCCGCCAAGCCCACGGAGCAGCGGCCGG

TACAGGTGGCTAAGCGATTGGAGAAGTTCAAAACAACAATTTTCACACAG

ATGAGCATGCTCGCAGTGAAGCATGGAGCAATAAACCTTGGACAGGGGTT

TCCCAATTTTGATGGCCCTGACTTTGTCAAAGATGCTGCTATTGAGGCTA

TCAAAGCTGGAAAGAATCAGTATGCAAGAGGATATGGTGTGCCTGAATTG

AACTCAGCTGTTGCTGAGAGATTTCTCAAGGACAGTGGATTGCACATCGA

TCCTGATAAGGAAGTTACTGTTACATCTGGGTGCACAGAAGCAATAGCTG

CAACGATATTGGGTCTGATCAACCCTGGGGATGAAGTCATACTGTTTGCT

CCATTCTATGATTCTTATGAGGCTACACTGTCCATGGCTGGTGCGAATGT

CAAAGCCATTACACTCCGCCCTCCGGACTTTGCAGTCCCTCTTGAAGAGC

TAAAGGCTGCAGTCTCGAAGAATACCAGAGCAATAATGATTAATACACCT

CACAACCCTACCGGGAAAATGTTCACAAGGGAGGAACTTGAGTTCATTGC

TGATCTCTGCAAGGAAAATGACGTGTTGCTCTTTGCCGATGAGGTCTACG

ACAAGCTGGCGTTTGAGGCGGATCACATATCAATGGCTTCTATTCCTGGC

ATGTATGAGAGGACCGTCACTATGAACTCCCTGGGGAAGACGTTCTCCTT

GACCGGATGGAAGATCGGCTGGGCGATAGCACCACCGCACCTGACATGGG

GCGTAAGGCAGGCACACTCCTTCCTCACATTCGCCACCTCCACGCCGATG

CAATCAGCAGCGGCGGCGGCCCTGAGAGCACCGGACAGCTACTTTGAGGA

GCTGAAGAGGGACTACGGCGCAAAGAAAGCGCTGCTGGTGGACGGGCTCA

AGGCGGCGGGCTTCATCGTCTACCCTTCGAGCGGAACCTACTTCATCATG

GTCGACCACACCCCGTTCGGGTTCGACAACGACGTCGAGTTCTGCGAGTA

CTTGATCCGCGAGGTCGGCGTCGTGGCCATCCCGCCAAGCGTGTTCTACC

TGAACCCGGAGGACGGGAAGAACCTGGTGAGGTTCACCTTCTGCAAGGAC

GACGACACGCTAAGGGCGGCGGTGGACAGGATGAAGGCCAAGCTCAGGAA

GAAATGATTGAGGGGCG

CACGTGTGA

SEQ ID NO: 22

Cambia 1201 + *Arabidopsis* GPT sequence (35S promoter from CaMV in italics)

CATGGAGTCAAAGATTCAAATAGAGGACCTAACAGAACTCGCCGTAAAGA

CTGGCGAACAGTTCATACAGAGTCTCTTACGACTCAATGACAAGAAGAAA

ATCTTCGTCAACATGGTGGAGCACGACACACTTGTCTACTCCAAAAATAT

CAAAGATACAGTCTCAGAAGACCAAAGGGCAATTGAGACTTTTCAACAAA

GGGTAATATCCGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCAC

TTTATTGTGAAGATAGTGGAAAAGGAAGGTGGCTCCTACAAATGCCATCA

TTGCGATAAAGGAAAGGCCATCGTTGAAGATGCCTCTGCCGACAGTGGTC

CCAAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTT

CCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACTGACGT

AAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATAT

AAGGAAGTTCATTTCATTTGGAGAGAACACGGGGGACTCTTGACC ATGTA

CCTGGACATAAATGGTGTGATGATCAAACAGTTTAGCTTCAAAGCCTCTC

TTCTCCCATTCTCTTCTAATTTCCGACAAAGCTCCGCCAAAATCCATCGT

CCTATCGGAGCCACCATGACCACAGTTTCGACTCAGAACGAGTCTACTCA

AAAACCCGTCCAGGTGGCGAAGAGATTAGAGAAGTTCAAGACTACTATTT

TCACTCAAATGAGCATATTGGCAGTTAAACATGGAGCGATCAATTTAGGC

CAAGGCTTTCCCAATTTCGACGGTCCTGATTTTGTTAAAGAAGCTGCGAT

CCAAGCTATTAAAGATGGTAAAAACCAGTATGCTCGTGGATACGGCATTC

CTCAGCTCAACTCTGCTATAGCTGCGCGGTTTCGTGAAGATACGGGTCTT

GTTGTTGATCCTGAGAAAGAAGTTACTGTTACATCTGGTTGCACAGAAGC

CATAGCTGCAGCTATGTTGGGTTTAATAAACCCTGGTGATGAAGTCATTC

TCTTTGCACCGTTTTATGATTCCTATGAAGCAACACTCTCTATGGCTGGT

GCTAAAGTAAAAGGAATCACTTTACGTCCACCGGACTTCTCCATCCCTTT

GGAAGAGCTTAAAGCTGCGGTAACTAACAAGACTCGAGCCATCCTTATGA

ACACTCCGCACAACCCGACCGGGAAGATGTTCACTAGGGAGGAGCTTGAA

ACCATTGCATCTCTCTGCATTGAAAACGATGTGCTTGTGTTCTCGGATGA

AGTATACGATAAGCTTGCGTTTGAAATGGATCACATTTCTATAGCTTCTC

TTCCCGGTATGTATGAAAGAACTGTGACCATGAATTCCCTGGGAAAGACT

TTCTCTTTAACCGGATGGAAGATCGGCTGGGCGATTGCGCCGCCTCATCT

GACTTGGGGAGTTCGACAAGCACACTCTTACCTCACATTCGCCACATCAA

CACCAGCACAATGGGCAGCCGTTGCAGCTCTCAAGGCACCAGAGTCTTAC

TTCAAAGAGCTGAAAAGAGATTACAATGTGAAAAAGGAGACTCTGGTTAA

GGGTTTGAAGGAAGTCGGATTTACAGTGTTCCCATCGAGCGGGACTTACT

TTGTGGTTGCTGATCACACTCCATTTGGAATGGAGAACGATGTTGCTTTC

TGTGAGTATCTTATTGAAGAAGTTGGGGTCGTTGCGATCCCAACGAGCGT

CTTTTATCTGAATCCAGAAGAAGGGAAGAATTTGGTTAGGTTTGCGTTCT

GTAAAGACGAAGAGACGTTGCGTGGTGCAATTGAGAGGATGAAGCAGAAG

CTTAAGAGAAAAGTCTGA

SEQ ID NO: 23

Cambia p1305.1 with (3′ end of) rbcS3C +

Arabidopsis GPT. Underlined ATG is start site, parentheses are the catI intron and the underlined

actagt is the speI cloning site used to splice in the *Arabidopsis* gene.

AAAAAAGAAAAAAAAAACATATCTTGTTTGTCAGTATGGGAAGTTTGAGA
TAAGGACGAGTGAGGGGTTAAAATTCAGTGGCCATTGATTTTGTAATGCC
AAGAACCACAAAATCCAATGGTTACCATTCCTGTAAGATGAGGTTTGCTA
ACTCTTTTTGTCCGTTAGATAGGAAGCCTTATCACTATATATACAAGGCG
TCCTAATAACCTCTTAGTAACCAATTATTTCAGCA

ATGG

TAGATCTGA
GG(GTAAATTTCTAGTTTTTCTCCTTCATTTTCTTGGTTAGGACCCTTTT
CTCTTTTTATTTTTTTGAGCTTTGATCTTTCTTTAAACTGATCTATTTTT
TAATTGATTGGTTATGGTGTAAATATTACATAGCTTTAACTGATAATCTG
ATTACTTTATTTCGTGTGTCTATGATGATGATGATAGTTACAG)AACCGA
CGA

ACTAGT

ATGTACCTGGACATAAATGGTGTGATGATCAAACAGTTTAC

CTTCAAAGCCTCTCTTCTCCCATTCTCTTCTAATTTCCGACAAAGCTCCG
CCAAAATCCATCGTCCTATCGGAGCCACCATGACCACAGTTTCGACTCAG
AACGAGTCTACTCAAAAACCCGTCCAGGTGGCGAAGAGATTAGAGAAGTT
CAAGACTACTATTTTCACTCAAATGAGCATATTGGCAGTTAAACATGGAG
CGATCAATTTAGGCCAAGGCTTTCCCAATTTCGACGGTCCTGATTTTGTT
AAAGAAGCTGCGATCCAAGCTATTAAAGATGGTAAAAACCAGTATGCTCG
TGGATACGGCATTCCTCAGCTCAACTCTGCTATAGCTGCGCGGTTTCGTG
AAGATACGGGTCTTGTTGTTGATCCTGAGAAAGAAGTTACTGTTACATCT
GGTTGCACAGAAGCCATAGCTGCAGCTATGTTGGGTTTAATAAACCCTGG
TGATGAAGTCATTCTCTTTGCACCGTTTTATGATTCCTATGAAGCAACAC
TCTCTATGGCTGGTGCTAAAGTAAAAGGAATCACTTTACGTCCACCGGAC

TTCTCCATCCCTTTGGAAGAGCTTAAAGCTGCGGTAACTAACAAGACTCG

AGCCATCCTTATGAACACTCCGCACAACCCGACCGGGAAGATGTTCACTA

GGGAGGAGCTTGAAACCATTGCATCTCTCTGCATTGAAAACGATGTGCTT

GTGTTCTCGGATGAAGTATACGATAAGCTTGCGTTTGAAATGGATCACAT

TTCTATAGCTTCTCTTCCCGGTATGTATGAAAGAACTGTGACCATGAATT

CCCTGGGAAAGACTTTCTCTTTAACCGGATGGAAGATCGGCTGGGCGATT

GCGCCGCCTCATCTGACTTGGGGAGTTCGACAAGCACACTCTTACCTCAC

ATTCGCCACATCAACACCAGCACAATGGGCAGCCGTTGCAGCTCTCAAGG

CACCAGAGTCTTACTTCAAAGAGCTGAAAAGAGATTACAATGTGAAAAAG

GAGACTCTGGTTAAGGGTTTGAAGGAAGTCGGATTTACAGTGTTCCCATC

GAGCGGGACTTACTTTGTGGTTGCTGATCACACTCCATTTGGAATGGAGA

ACGATGTTGCTTTCTGTGAGTATCTTATTGAAGAAGTTGGGGTCGTTGCG

ATCCCAACGAGCGTCTTTTATCTGAATCCAGAAGAAGGGAAGAATTTGGT

TAGGTTTGCGTTCTGTAAAGACGAAGAGACGTTGCGTGGTGCAATTGAGA

GGATGAAGCAGAAGCTTAAGAGAAAAGTCTGA

SEQ ID NO: 24

Arabidpsis GPT coding sequence (mature protein, no targeting sequence)

GTGGCGAAGAGATTAGAGAAGTTCAAGACTACTATTTTCACTCAAATGAG

CATATTGGCAGTTAAACATGGAGCGATCAATTTAGGCCAAGGCTTTCCCA

ATTTCGACGGTCCTGATTTTGTTAAAGAAGCTGCGATCCAAGCTATTAAA

GATGGTAAAAACCAGTATGCTCGTGGATACGGCATTCCTCAGCTCAACTC

TGCTATAGCTGCGCGGTTTCGTGAAGATACGGGTCTTGTTGTTGATCCTG

AGAAAGAAGTTACTGTTACATCTGGTTGCACAGAAGCCATAGCTGCAGCT

ATGTTGGGTTTAATAAACCCTGGTGATGAAGTCATTCTCTTTGCACCGTT

TTATGATTCCTATGAAGCAACACTCTCTATGGCTGGTGCTAAAGTAAAAG

GAATCACTTTACGTCCACCGGACTTCTCCATCCCTTTGGAAGAGCTTAAA

GCTGCGGTAACTAACAAGACTCGAGCCATCCTTATGAACACTCCGCACAA

CCCGACCGGGAAGATGTTCACTAGGGAGGAGCTTGAAACCATTGCATCTC

TCTGCATTGAAAACGATGTGCTTGTGTTCTCGGATGAAGTATACGATAAG

CTTGCGTTTGAAATGGATCACATTTCTATAGCTTCTCTTCCCGGTATGTA

TGAAAGAACTGTGACCATGAATTCCCTGGGAAAGACTTTCTCTTTAACCG

GATGGAAGATCGGCTGGGCGATTGCGCCGCCTCATCTGACTTGGGGAGTT

CGACAAGCACACTCTTACCTCACATTCGCCACATCAACACCAGCACAATG

GGCAGCCGTTGCAGCTCTCAAGGCACCAGAGTCTTACTTCAAAGAGCTGA

AAAGAGATTACAATGTGAAAAAGGAGACTCTGGTTAAGGGTTTGAAGGAA

GTCGGATTTACAGTGTTCCCATCGAGCGGGACTTACTTTGTGGTTGCTGA

TCACACTCCATTTGGAATGGAGAACGATGTTGCTTTCTGTGAGTATCTTA

TTGAAGAAGTTGGGGTCGTTGCGATCCCAACGAGCGTCTTTTATCTGAAT

CCAGAAGAAGGGAAGAATTTGGTTAGGTTTGCGTTCTGTAAAGACGAAGA

GACGTTGCGTGGTGCAATTGAGAGGATGAAGCAGAAGCTTAAGAGAAAAG

TCTGA

SEQ ID NO: 25

Arabidpsis GPT amino acid sequence (mature protein, no targeting sequence)

VAKRLEKFKTTIFTQMSILAVKHGAINLGQGFPNFDGPDFVKEAAIQAIK

DGKNQYARGYGIPQLNSAIAARFREDTGLVVDPEKEVTVTSGCTEAIAAA

MLGLINPGDEVILFAPFYDSYEATLSMAGAKVKGITLRPPDFSIPLEELK

AAVTNKTRAILMNTPHNPTGKMFTREELETIASLCIENDVLVFSDEVYDK

LAFEMDHISIASLPGMYERTVTMNSLGKTFSLTGWKIGWAIAPPHLTWGV

RQAHSYLTFATSTPAQWAAVAALKAPESYFKELKRDYNVKKETLVKGLKE

VGFTVFPSSGTYFVVADHTPFGMENDVAFCEYLIEEVGVVAIPTSVFYLN

PEEGKNLVRFAFCKDEETLRGAIERMKQKLKRKV

SEQ ID NO: 26

Grape GPT amino acid sequence (mature protein, no targeting sequence)

VAKRLEKFKTTIFTQMSMLAIKI:IGAINLGQGFPNFDGPEFVKEAAIQA

IKDGKNQYARGYGVPDLNSAVADRFKKDTGLWDPEKEVTVTSGCTEAIAA

TMLGLINPGDEVILFAPFYDSYEATLSMAGAQIKSITLRPPDFAVPMDEL

KSAISKNTRAILINTPHNPTGKMFTREELNVIASLCIENDVLVFTDEVYD

KLAFEMDHISMASLPGMYERTVTMNSLGKTFSLTGWKIGVVTVAPPHLTW

GVRQAHSFLTFATCTPMQWAAATALRAPDSYYEELKRDYSAKKAILVEGL

KAVGFRVYPSSGTYFVVVDHTPFGLKDDIAFCEYLIKEVGVVAIPTSVFY

LHPEDGKNLVRFTFCKDEGTLRAAVERMKEKLKPKQ

SEQ ID NO: 27

Rice GPT amino acid sequence (mature protein, no targeting sequence)

VAKRLEKFKTTIFTQMSMLAIKHGAINLGQGFPNFDGPDFVKEAAIQAIN

AGKNQYARGYGVPELNSAIAERFLKDSGLQVDPEKEVTVTSGCTEAIAAT

ILGLINPGDEVILFAPFYDSYEATLSMAGANVKAITLRPPDFSVPLEELK

AAVSKNTRAIMINTPHNPTGKMFTREELEFIATLCKENDVLLFADEVYDK

LAFEADHISMASIPGMYERTVTMNSLGKTFSLTGWKIGWAIAPPHLTWGV

RQAHSFLTFATCTPMQAAAAAALRAPDSYYEELRRDYGAKKALLVNGLKD

AGFIVYPSSGTYFVMVDHTPFGFDNDIEFCEYLIREVGVVAIPPSVFYLN

PEDGKNLVRFTFCKDDETLRAAVERMKTKLRKK

SEQ ID NO: 28

Soybean GPT amino acid sequence (−1 mature protein, no targeting sequence)

AKRLEKFQTTIFTQMSLLAIKHGAINLGQGFPNFDGPEFVKEAAIQAIRD

GKNQYARGYGVPDLNIAIAERFKKDTGLVVDPEKEITVTSGCTEAIAATM

IGLINPGDEVIMFAPFYDSYEATLSMAGAKVKGITLRPPDFAVPLEELKS

TISKNTRAILINTPHNPTGKMFTREELNCIASLCIENDVLVFTDEVYDKL

AFDMEHISMASLPGMFERTVTLNSLGKTFSLTGWKIGWAIAPPHLSWGVR

QAHAFLTFATAHPFQCAAAAALRAPDSYYVELKRDYMAKRAILIEGLKAV

GFKVFPSSGTYFVVVDHTPFGLENDVAFCEYLVKEVGVVAIPTSVFYLNP

EEGKNLVRFTFCKDEETIRSAVERMKAKLRKVD

SEQ ID NO: 29

Barley GPT amino acid sequence (mature protein, no targeting sequence)

VAKRLEKFKTTIFTQMSMLAVKHGAINLGQGFPNFDGPDFVKDAAIEAIK
AGKNQYARGYGVPELNSAVAERFLKDSGLHIDPDKEVTVTSGCTEAIAAT
ILGLINPGDEVILFAPFYDSYEATLSMAGANVKAITLRPPDFAVPLEELK
AAVSKNTRAIMINTPHNPTGKMFTREELEFIADLCKENDVLLFADEVYDK
LAFEADHISMASIPGMYERTVTMNSLGKTFSLTGWKIGWAIAPPHLTWGV
RQAHSFLTFATSTPMQSAAAAALRAPDSYFEELKRDYGAKKALLVDGLKA
AGFIVYPSSGTYFIMVDHTPFGFDNDVEFCEYLIREVGVVAIPPSVFYLN
PEDGKNLVRFTFCKDDDTLRAAVDRMKAKLRKK

SEQ ID NO: 30

Zebra fish GPT amino acid sequence (mature protein, no targeting sequence)

VAKRLEKFKTTIFTQMSMLAIKHGAINLGQGFPNFDGPDFVKEAAIQAIR
DGNNQYARGYGVPDLNIAISERYKKDTGLAVDPEKEITVTSGCTEAIAAT
VLGLINPGDEVIVFAPFYDSYEATLSMAGAKVKGITLRPPDFALPIEELK
STISKNTRAILLNTPHNPTGKMFTPEELNTIASLCIENDVLVFSDEVYDK
LAFDMEHISIASLPGMFERTVTMNSLGKTFSLTGWKIGWAIAPPHLTWGV
RQAHAFLTFATSNPMQWAAAVALRAPDSYYTELKRDYMAKRSILVEGLKA
VGFKVFPSSGTYFVWDHTPFGHENDIAFCEYLVKEVGVVAIPTSVFYLNP
EEGKNLVRFTFCKDEGTLRAAVDRMKEKLRK

SEQ ID NO: 31

Bamboo GPT amino acid sequence (mature protein, no targeting sequence)

VAKRLEKFKTTIFTQMSMLAIKHGAINLGQGFPNFDGPDFVKEAAIQAIN
AGKNQYARGYGVPELNSAVAERFLKDSGLQVDPEKEVTVTSGCTEAIAAT
ILGLINPGDEVILFAPFYDSYEATLSMAGANVKAITLRPPDFAVPLEELK

ATVSKNTRAIMINTPHNPTGKMFSREELEFIATLCKKNDVLLFADEVYDK

LAFEADHISMASIPGMYERTVTMNSLGKTFSLTGWKIGWAIAPPHLTWGV

RQAHSFLTFATCTPMQSAAAAALRAPDSYYGELKRDYGAKKAILVDGLKA

AGFIVYPSSGTYFVMVDHTPFGFDNDIEFCEYLIREVGWAIPPSVFYLNP

EDGKNLVRFTFCKDDDTLRAAVERMKTKLRKK

CLASSIFICATIONS

U.S. Classification	800/290, 800/298, 800/278
International Classification	A01H5/00, C12N15/82
Cooperative Classification	C12N9/1096, C12N15/8261
European Classification	C12N15/82C8, C12N9/10F

LEGAL EVENTS

Date	Code	Event	Description
Mar 25, 2010	AS	Assignment	**Owner name:** LOS ALAMOS NATIONAL SECURITY, LLC, NEW MEXICO **Free format text:** ASSIGNMENT OF ASSIGNORS INTEREST;ASSIGNORS:UNKEFER, PAT J.;ANDERSON, PENELOPE S.;REEL/FRAME:024142/0132 **Effective date:** 20091216
Jan 13, 2010	AS	Assignment	**Owner name:** MAINE SYSTEM BOARD OF TRUSTEES, UNIVERSITY OF, MAI **Effective date:** 20100113 **Free format text:** ASSIGNMENT OF ASSIGNORS INTEREST;ASSIGNOR:KNIGHT, THOMAS J., DR.;REEL/FRAME:023777/0416
Nov 5, 2009	AS	Assignment	**Free format text:** CONFIRMATORY LICENSE;ASSIGNOR:LOS ALAMOS NATIONAL SECURITY;REEL/FRAME:023472/0933 **Effective date:** 20090915 **Owner name:** U.S. DEPARTMENT OF ENERGY, DISTRICT OF COLUMBIA